

Propostas de Aplicações para Tratamento Lingüístico Computacional do Esperanto*

Moisés Braga Ribeiro
moises.ribeiro@ic.unicamp.br

Data: 2 de julho de 2001

Resumo

Este artigo aborda a língua planejada Esperanto e propõe aplicações computacionais para seu tratamento lingüístico. Comparações com outras línguas (tanto naturais quanto planejadas) são feitas de modo a demonstrar a regularidade do Esperanto bem como a possibilidade de seu tratamento lingüístico computacional ser mais simples. As aplicações propostas não são dessa maneira para processamento lingüístico em geral —são para processamento específico de Esperanto, tirando proveito de sua gramática extensivamente regular para simplificar e direcionar suas ações. Tais aplicações envolvem análises lingüísticas nos níveis fonético, morfológico e sintático e outras aplicações que utilizam estas análises como ponto de partida, como interfaces textuais, procura “inteligente” de texto e tradução automática. Um conhecimento prévio de Esperanto não é necessário para a plena compreensão deste artigo.

Palavras-chave: Esperanto, processamento de linguagem natural, análise morfológica, análise sintática.

*Este artigo é o trabalho final da disciplina MO622 (Fatores Humanos em Sistemas de Informação), ministrada pela professora doutora Maria Cecília Calani Baranauskas e realizada no primeiro semestre de 2001, no Instituto de Computação da Universidade Estadual de Campinas (IC/UNICAMP).

Sumário

1	Introdução	3
2	Histórico do Esperanto	3
2.1	Projeto	4
2.2	Literatura	5
2.3	Movimento	5
3	Comparação do Esperanto com Outras Línguas	6
3.1	Comparação com Línguas Planejadas	6
3.1.1	Pasigrafias	7
3.1.2	Línguas Planejadas <i>a Priori</i>	8
3.1.3	Línguas Planejadas <i>a Posteriori</i>	8
3.1.4	Línguas Planejadas Mistas	10
3.2	Comparação com Línguas Naturais	12
3.2.1	Ortografia	13
3.2.2	Formação de Palavras	14
3.2.3	Construção de Orações	16
4	Resumo Gramatical do Esperanto	18
4.1	Morfologia	18
4.2	Sintaxe	19
4.2.1	Papéis Sintáticos dos Sintagmas	20
4.2.2	Observações sobre a Construção de Sintagmas	21
5	Tratamento Computacional do Esperanto	22
5.1	Análises Lingüísticas Básicas	23
5.1.1	Análise Fonética	23
5.1.2	Análise Morfológica	23
5.1.3	Análise Sintática	25
5.1.4	Relações entre as Análises	27
5.2	Aplicações das Análises Lingüísticas	28
5.2.1	Procura “Inteligente” de Texto	28
5.2.2	Análise Enfática	28
5.2.3	Síntese de Voz	28
5.2.4	Análise Semântica	28
5.2.5	Interfaces Textuais	29
5.2.6	Construção de Dicionários Bilíngües	29
5.2.7	Resolução de Referências	30
5.2.8	Tradução Automática para Línguas Naturais	30
6	Conclusão	31
A	Termos Lingüísticos	32
B	Alfabeto Esperantista	34

1 Introdução

Este artigo trata da língua planejada Esperanto, e propõe aplicações computacionais para tratamento de textos que podem tirar proveito de sua gramática regular. Comparações entre o Esperanto e línguas planejadas e naturais são apresentadas, com o intuito de ressaltar as qualidades lingüísticas do Esperanto, que se tornam vantajosas durante o processamento computacional. A comparação com línguas planejadas leva em conta aspectos utilizados quando da criação da língua, mostrando os tipos de línguas planejadas existentes. A comparação com línguas naturais coloca-as no papel de línguas estrangeiras ou segundas línguas, destacando a facilidade de aprendizado do Esperanto, devido à sua regularidade. A comparação é realizada nos níveis fonético, morfológico e sintático.

Não é preciso saber Esperanto para entender o conteúdo deste artigo. Sua gramática é explicada e comparada às de outras línguas sempre com exemplos elucidativos, juntamente com as traduções e outras informações necessárias. Não são considerados aspectos motivacionais (por que aprender Esperanto), embora o texto também tenha o intuito de despertar a curiosidade do leitor. O principal objetivo é avaliar possibilidades de processamento computacional do Esperanto —um campo ainda quase inexplorado.

As seguintes convenções são utilizadas no decorrer do texto:

- [som] — representa um som (ou uma seqüência de sons) denotado, se não indicado de outra maneira, pelo alfabeto fonético internacional [IPA]; por exemplo, os sons correspondentes à letra xis nas palavras *xícara*, *exame* e *explorar* são representados respectivamente por [f], [z] e [s];
- *errado — representa uma palavra hipotética, no entanto formada de uma maneira permitida pela gramática;
- destaque — destaca um morfema (ou mesmo letras soltas) dentro de uma palavra;
- primitiva → derivada — representa a derivação de uma palavra.

A organização deste artigo é a seguinte: a Seção 2 apresenta um breve histórico do Esperanto; a Seção 3 mostra comparações entre o Esperanto e algumas línguas planejadas e naturais; a Seção 4 resume pontos importantes da gramática do Esperanto discutidos em outras partes; a Seção 5 discorre sobre análises lingüísticas computacionais aplicadas ao Esperanto e lista oito aplicações para elas; a Seção 6 apresenta as considerações finais. No Apêndice A aparecem definições para os termos lingüísticos utilizados, e no Apêndice B esclarecimentos sobre a pronúncia de todas as letras usadas no alfabeto esperantista.

2 Histórico do Esperanto

Dentre todos os projetos de línguas planejadas, somente o Esperanto, com seus mais de 100 anos de história, chegou ao último nível que um projeto como este

pode alcançar: ao uso como língua falada e escrita em todos os sentidos (inclusive poesia, literatura, jargões técnicos, gíria e linguagem chula) por um grupo relativamente grande e organizado internacionalmente, que realiza congressos periódicos e publica material de assuntos variados [Blan89]. O Esperanto era um projeto há cem anos, mas hoje em dia pode ser considerado como virtualmente equivalente a uma língua natural, pois é sujeito aos mesmos fenômenos lingüísticos que fazem as línguas naturais evoluir em termos de vocabulário e estilo [Piro89].

A seguir se encontram breves históricos do Esperanto enquanto projeto (Seção 2.1), meio para a produção de literatura (Seção 2.2) e movimento social (Seção 2.3).

2.1 Projeto

O autor do Esperanto Ludwik Lejzer Zamenhof (1859–1917) cresceu num ambiente com diversas línguas —russo, alemão, polonês e iídiche— no qual cada grupo lingüístico hostilizava os outros. Com quinze anos já sabia também francês, grego e latim, além destas quatro línguas. Desde cedo percebeu a necessidade de uma maneira comum de expressão para todos os povos como meio de aproximação e entendimento. A grande diferença de Zamenhof como criador de um projeto de língua planejada foi a de reconhecer os caracteres social e evolutivo das línguas [Jant93]. Um projeto como este só consegue se tornar uma língua de verdade através de uma comunidade que o usa, apóia e divulga. A estrutura lingüística do projeto não pode impedir a evolução natural da língua pela assimilação de mais palavras ou por mudanças de estilo.

Os principais livros que embasaram o Esperanto enquanto projeto aparecem na linha do tempo abaixo.

- 1887** *Lingvo Internacia* (Língua Internacional), primeira obra de Zamenhof, sob o pseudônimo *Doktoro Esperanto* (doutor que tem esperança), aos 28 anos. Foi publicada primeiramente em russo, e em seguida em polonês, francês, alemão e inglês. O pseudônimo de Zamenhof acabou se tornando o próprio nome da língua.
- 1888** *Dua Libro de l' Lingvo Internacia* (Segundo Livro da Língua Internacional).
- 1889** Suplemento do segundo livro. Esta foi a última palavra de Zamenhof no papel de autor da língua. A partir de então ele se considerou um esperantista comum, sem privilégios.
- 1894** *Universala Vortaro* (Dicionário Universal), com traduções em francês, inglês, alemão, russo e polonês.
- 1903** *Fundamenta Krestomatio de la Lingvo Esperanto* (Coletânea Fundamental da Língua Esperanto), antologia com exercícios, artigos, palestras, histórias, poesia e prosa, em textos originais e traduzidos.

1905 *Fundamento de Esperanto*, com as 16 regras gramaticais fundamentais, exercícios e o dicionário de 1894. Neste ano também ocorreu o primeiro congresso mundial, em Boulogne-sur-Mer, na França.

2.2 Literatura

A literatura esperantista já é bastante vasta. Grandes coleções podem ser encontradas no escritório principal da Associação Universal de Esperanto em Roterdã (Holanda), na Associação Britânica de Esperanto em Londres, na biblioteca municipal de La Chaux-de-Fonds na Suíça e no Museu Internacional do Esperanto, uma parte da biblioteca nacional austríaca em Viena [Schu89a]. Coleções de portes variados existem em cada um das centenas de grupos esperantistas espalhados pelo mundo.

Obras importantes da literatura mundial já foram traduzidas para o Esperanto, como livros de Dante, Camões, Shakespeare, Goethe, Sartre, Machado de Assis e muitos outros autores, além de textos religiosos como o Bhagavad-Gita, a Bíblia, o Alcorão e o Livro dos Espíritos [Pass93]. Na prática a vantagem do Esperanto está no fato que os tradutores geralmente trabalham vertendo de sua língua materna para o Esperanto, contrapondo-se assim ao procedimento usual de vir de uma língua estrangeira (mesmo que bem compreendida) para a sua língua materna.

A literatura em Esperanto pode ser separada em três períodos [Jant93]: o primeiro, desde a publicação do projeto até o início da primeira Guerra Mundial; o segundo, durante e entre as Guerras Mundiais; e o terceiro, do pós-guerra até atualmente. O primeiro período é caracterizado pelo embasamento da língua, no qual os escritores são principalmente tradutores, teóricos e propagandistas. No segundo período o estudo lingüístico sobre o Esperanto se desenvolve, ao mesmo tempo em que a literatura chega à maturidade, com autores cuja atividade se estende por décadas. O terceiro período é marcado pela proliferação de jornais e publicações (recentemente também via internet) e pela chegada do Esperanto aos países do leste asiático.

2.3 Movimento

Desde o começo do movimento os esperantistas formaram associações locais, sendo a mais antiga a de Nuremberg, na Alemanha (1888). Existem atualmente associações nacionais em 117 países, além de dezenas de associações internacionais científicas, ideológicas e de atividades de lazer. A associação internacional mais importante é a *Universala Esperanto Asocio* (Associação Universal de Esperanto), fundada em 1908, com sede em Roterdã. Ela promove o congresso mundial, que tem reunido centenas de pessoas anualmente em diversas partes do mundo.¹

O movimento esperantista edita publicações científicas, literárias, culturais, políticas, religiosas e filosóficas, além de outras relativas ao próprio Esperanto.

¹O congresso mundial de 2002 ocorrerá no Brasil, em Fortaleza.

Também existem revistas nacionais planejadas para informar esperantistas sobre determinado país, algumas com apoio governamental.

Com relação à educação do Esperanto, estudos comprovam que ele é a língua mais rápida de ser aprendida e a língua na qual o estudante médio pode esperar alcançar o nível mais alto de proficiência [Jant93]. Já em 1933 um estudo entre universitários americanos avaliou que a proficiência adquirida em um ano de Esperanto equivalia a de quatro anos de francês ou alemão. Nos anos 70 uma pesquisa mostrou que o nível lingüístico de crianças francesas era equivalente depois de 1500 horas de estudo de inglês e apenas 150 horas de Esperanto.

Além de ser facilmente aprendido, o Esperanto ainda contribui para que a pessoa aprenda mais sobre sua própria língua e sobre outras (por meio de comparações com a gramática regular), e também para aumentar sua motivação em aprender sobre outros países. A cultura esperantista é bem descrita neste trecho de [Jant93]:

“A cultura que o Esperanto carrega com ele deriva de um único fato social e psicológico fundamental: quando milhares de pessoas de diferentes línguas e culturas decidem livremente falar a mesma língua, emerge um novo tipo de sociedade, talvez ainda embrionária, um ‘povo esperantista’ com uma identidade especial. Ao mesmo tempo em que não esquece ou coloca de lado sua cultura nativa, é consciente de outras culturas e em conseqüência compreende a relatividade da sua própria. Embora os esperantistas difiram entre si em opiniões filosóficas, políticas, artísticas e outras (...), mesmo assim o Esperanto lhes permite a oportunidade de se informarem diretamente e de fazer contatos interpessoais, e isso cria um forte senso de tolerância. Por esta razão os ditadores do século 20 temeram, corretamente, que a língua pudesse emancipar as pessoas dos dogmas oficiais, e a proibiram.”

3 Comparação do Esperanto com Outras Línguas

Nesta seção o Esperanto será comparado a algumas línguas, tanto planejadas (Seção 3.1) quanto naturais (Seção 3.2). As vantagens de sua gramática são ressaltadas nos pontos-de-vista lingüístico e computacional.

3.1 Comparação com Línguas Planejadas

No decorrer deste artigo será usado o termo *língua planejada* para denominar línguas criadas conscientemente, de forma original ou por meio de “simplificações” de línguas naturais.² Outros termos são usados para descrevê-las [Blan89], dentre os quais *língua artificial*, *língua universal*, *língua internacional*. O primeiro

²O termo *língua natural* é usado no sentido de língua de um grupo étnico, uma região, um país ou um povo. É um termo lingüístico clássico, atualmente substituído por *língua étnica*, mas ainda usado no contexto da Ciência de Computação.

tem um certo sentido pejorativo quando aplicado a línguas, dando a entender que a língua não foi feita para pessoas e sim para máquinas. Os termos *universal* e *internacional* podem até mesmo levar a mal-entendidos sobre a função da língua. Uma língua universal é para ser usada no mundo inteiro, substituindo as outras? Uma língua que vários países usam já não é internacional?

O termo *língua planejada* é neutro nestes dois pontos, pois indica apenas que um planejamento foi feito quando da criação da língua. Algumas estão em pleno uso atualmente, como o Hebraico moderno, usado em Israel, e o Bahasa Indonesia, na Indonésia. Embora adequado, este termo tem um significado bastante amplo, abarcando diversos tipos de línguas. Elas são classificadas normalmente de acordo com duas características principais [Blan89, Jant93]:

1. Intenção de uso
 - (a) pasigrafias: línguas voltadas somente para escrita;
 - (b) pasilalias: línguas para escrita e conversação;
2. Semelhança com línguas naturais
 - (a) *a priori*: línguas baseadas em classificações pré-concebidas de conceitos, palavras com raízes artificiais, derivação esquemática, fonética, morfologia e sintaxe regulares;
 - (b) *a posteriori*: línguas que usam como modelo línguas naturais, com léxico semelhante (em graus variados) e irregularidades gramaticais freqüentes.

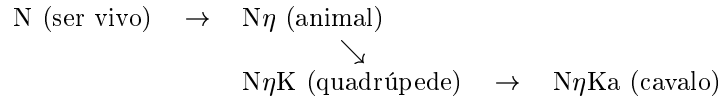
A seguir aparecem exemplos de algumas línguas planejadas classificadas da maneira indicada acima. As pasigrafias em geral são abordadas brevemente (Seção 3.1.1); as pasilalias estão separadas em *a priori* (Seção 3.1.2), *a posteriori* (Seção 3.1.3) e mistas (Seção 3.1.4). O Esperanto é brevemente comparado a duas línguas importantes de seu grupo (pasilalias mistas). As pasilalias correspondem à noção comum do termo *língua*, e desta maneira são denominadas abaixo.

3.1.1 Pasigrafias

Pasigrafias são sistemas puramente simbólicos (voltados apenas para escrita) que de alguma forma representam conceitos em símbolos, seja com letras, ideogramas ou até mesmo com notas musicais, como o *Solresol* de Jean Sudre (1866). As pasigrafias podem ser *a priori* ou *a posteriori*. A idéia central das pasigrafias *a priori* é o postulado cartesiano que diz que “a invenção da língua universal depende da verdadeira filosofia” [Jant93], que originou sistemas nos quais a classificação das noções em conjuntos é o mais importante.

Como exemplo de pasigrafia *a priori* pode ser citada a *Ars Signorum* de George Dalgarno (1661), que define 17 classes de idéias, cada qual representada por uma letra maiúscula, das quais todas as idéias relacionadas podem ser

derivadas através de combinações de letras gregas e latinas. Eis exemplos de derivações, citadas em [Jant93]:



Um exemplo de pasigrafia *a posteriori* é o *Safo* de André Eckardt (1952). Ele utiliza 180 símbolos básicos e um modelo parecido com a escrita chinesa para fazer relações entre os símbolos. Em [Blan89] aparecem os seguintes exemplos:

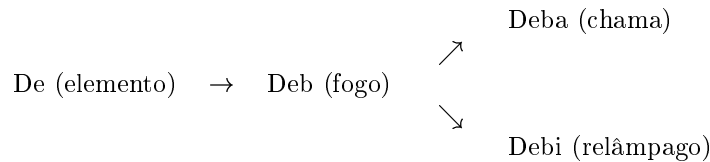


Embora algumas associações tenham surgido no século 19, hoje em dia nenhum sistema pasigráfico é usado extensivamente.

3.1.2 Línguas Planejadas *a Priori*

O mesmo postulado cartesiano citado na Seção 3.1.1 deu origem a línguas classificatórias que podem ser faladas. Estas línguas, embora tenham um léxico arbitrariamente inventado, não chegam a ser inovadoras num ponto de vista mais profundo, pois utilizam uma sintaxe baseada nas línguas naturais europeias — simples e estritamente regular, no entanto [Jant93].

A formação arbitrária das palavras numa língua *a priori* é bem ilustrada pela “língua filosófica” de John Wilkins (1668). Eis exemplos de [Blan89]:



As classificações empregadas em línguas *a priori* têm uma certa semelhança com outros sistemas de classificação de conceitos, como o sistema decimal usado para classificar assuntos em bibliotecas.

3.1.3 Línguas Planejadas *a Posteriori*

É conveniente dividir o conjunto de línguas planejadas *a posteriori* de acordo com o número de línguas naturais envolvidas na criação. As que se baseiam em apenas uma *adaptam* esta língua natural de uma certa maneira, para que presumivelmente seu aprendizado seja mais fácil. Já as chamadas línguas *naturalísticas* recebem influências de um conjunto de línguas, normalmente relacionadas entre si, procurando ser “pontos médios” lexicais e gramaticais deste conjunto.

Línguas adaptadas Têm o objetivo de “simplificar” ou “reformular” línguas naturais atuais ou antigas. Dois exemplos importantes desta vertente são o *Latino sine flexione* de Giuseppe Peano (1904) e o *Basic English* de Charles Ogden (1931). Ambos procuram simplificar suas línguas de origem, mas o fazem de maneiras diferentes.

O Latino sine flexione retira do latim clássico o complicado sistema inflexional, que marca as funções sintáticas das palavras com desinências de caso. Eis a oração do pai-nosso em Latino sine flexione e em latim clássico [Chri01]:

Latino sine flexione	Latim clássico
Patre nostro, qui es in celos, que tuo nomine fi sanctificato. Que tuo regno adveni. Que tua voluntate es facta sicut in celo et in terra. Da hodie ad nos nostro pane quotidiano.	Pater noster qui es in coelis, sanctificetur nomem tuum. Adveniat regnum tuum. Fiat voluntas tua, sicut in caelo et in terra. Panem nostrum quotidianum da nobis hodie.
Et remitte ad nos nostros debitos, sicut et nos remitte ad nostros debitores.	Et dimitte nobis debita nostra, sicut et nos dimittimus debitoribus nostris.
Et non induce nos in tentatione, sed libera nos ab malo.	Et ne nos inducas in tentationem, sed libera nos a malo.

O Basic English utiliza apenas 850 palavras básicas do vocabulário do inglês. Como resultado muitas palavras têm que ser reescritas por meio de circumlocações, tornando a língua um tanto estranha para os falantes nativos. Eis alguns exemplos de circumlocações, citadas em [Blan89] e [Pass93]:

Basic English	Inglês
to come down (vir para baixo)	to descend (descer)
to go down (ir para baixo)	
to go from place to place without aim (ir de lugar para lugar sem objetivo)	to wander (zanzar)
I have thoughts in my mind (eu tenho pensamentos em minha mente)	I think (eu penso)

Estas duas línguas revelam as duas principais falhas destes tipos de línguas planejadas [Jant93]: a língua original pode ficar um tanto distorcida (falha no aspecto estético) ou o vocabulário pode ficar muito restrito (falha na expressividade). No entanto ambas chegaram a ser utilizadas durante algum tempo em correspondências internacionais e em textos literários e científicos.

Línguas naturalísticas Têm a pretensão de manter a “naturalidade” e a “beleza” das línguas naturais (latinas), com o intuito de serem de aprendizado mais fácil. No entanto aproximam-se demais das línguas naturais, contando também com irregularidades e arbitrariedades [Jant93], como letras diferentes

para um mesmo som, sufixos diferentes para representar a mesma idéia, alomorfa das raízes, etc.

O resultado final acaba sendo o oposto: a língua tem suas palavras rapidamente reconhecidas, o que torna sua leitura fácil e sem esforço, mas torna-se difícil de memorizar e escrever corretamente, pois as palavras devem ser aprendidas como numa língua natural —tanto palavras primitivas quanto derivadas devem ser decoradas. Isto fez com que essas línguas não conquistassem uma comunidade grande durante muito tempo.

Exemplos representativos desta categoria são as línguas *Occidental*, de Edgar de Wahl (1922), e *Interlingua*, de Alexander Gode (1951). Os exemplos abaixo em Occidental [Pass93] mostram adjetivos derivados de substantivos com um mesmo final, que no entanto apresentam sufixos inteiramente diferentes, como ocorre em português.

compensation → compensatori (compensação → compensatório)

intention → intentional (intenção → intencional)

illustration → illustrativ (ilustração → ilustrativo)

reation → reactionari (reação → reacionário)

Na Interlingua o fenômeno da alomorfa é freqüente, devido à sua influência latina. Os exemplos abaixo [Pass93] mostram que neste ponto esta língua é semelhante ao português:

distinguer → distinction (distinguir → distinção)

corriger → correction (corrigir → correção)

cognoscer → cognition (conhecer → conhecimento)

ager → action (agir → ação)

3.1.4 Línguas Planejadas Mistas

Alguns projetos de línguas planejadas são melhor classificados como mistos, pois combinam esquematização e regularidade com raízes provenientes de línguas naturais. O Esperanto encontra-se nesta categoria, uma vez que utiliza raízes majoritariamente latinas e uma gramática totalmente esquematizada e regular.

As línguas mistas podem pender para o lado *a priori*, como o *Volapük* de Johann Schleyer (1880), ou para o lado *a posteriori*, como o *Ido* de Louis Couturat e Louis de Beaufront (1907). O Esperanto se encontra numa posição média, que fica constatada com a comparação deste com essas duas línguas.

Volapük A primeira língua planejada a reunir milhares de adeptos e a ter jornais e livros publicados por uma década. Três congressos internacionais foram realizados, que acabaram por revelar suas deficiências. A intolerância de seu autor em acatar reformas acabou causando divisões no movimento volapukuista e depois o seu colapso.

Esta língua possui um alfabeto fonético e uma gramática regular, como o Esperanto, mas simplifica as raízes provenientes de línguas naturais de uma maneira que às vezes as torna quase irreconhecíveis. Eis alguns exemplos citados em [Pass93], com a tradução em Esperanto em itálico:

Yulop (Europa, <i>Eŭropo</i>)	min (animal, <i>besto</i>)
Melop (América, <i>Ameriko</i>)	fem (fermentar, <i>fermenti</i>)
Bodugän (Portugal, <i>Portugaliao</i>)	fom (forma, <i>formo</i>)
Basil (Brasil, <i>Brazilo</i>)	mab (mármore, <i>marmoro</i>)

Outra característica em comum com o Esperanto é o uso de afixos (prefixos e sufixos) para realizar a derivação de palavras. Somente o método é o mesmo, no entanto; o uso é bem diferente. A gramática do Volapük, apesar de regular, é bastante complicada. Os verbos, por exemplo, são extremamente compactos e incluem também um sufixo equivalente ao pronome pessoal, como demonstra o exemplo abaixo [Jant93]:³

<p style="text-align: center;">pilöfoböv (eu teria sido amado, <i>mi estus estinta amita</i>)</p>	{	<p>p = voz passiva i = tempo mais-que-perfeito löf = raiz do verbo amar ob = primeira pessoa do singular öv = modo condicional</p>
---	---	--

O Esperanto, em comparação, utiliza apenas seis desinências verbais que marcam tempo e modo. A tabela abaixo mostra as conjugações do verbo *kanti* (cantar). Como não há concordância de pessoa, um pronome foi incluído na tabela para facilitar a tradução.

Conjugação	Tradução	Tempo / modo
kant <u>i</u>	cantar	modo infinitivo
vi kant <u>as</u>	você canta	tempo presente do modo indicativo
vi kant <u>is</u>	você cantou	tempo passado do modo indicativo
vi kant <u>os</u>	você cantará	tempo futuro do modo indicativo
vi kant <u>u</u>	cante (você)	modo imperativo
vi kant <u>us</u>	você cantaria	modo imaginário

O modo imaginário é usado nos verbos que representam ações ou estados irrealis, imaginários ou fantasiosos [Wenn01], sendo semelhante (mas não totalmente equivalente) ao futuro do pretérito (ou condicional) do português.

Ido Uma das reformas propostas para deixar o Esperanto mais “natural”, isto é, parecido com o latim. O Ido incorporou algumas concessões das línguas naturalísticas ao Esperanto. Dois exemplos são o alfabeto não-fonético e o aparecimento de um verbo irregular, *esar* (ser).

³A forma verbal com três palavras presente na tradução em Esperanto não é usada na prática, embora seja possível. Normalmente a frase é escrita de outra maneira, de forma mais clara [Wenn01].

Embora alegassem que sua língua é mais simples que o Esperanto, os autores do Ido promoveram uma reforma no sistema de pronomes com vistas a separar claramente os gêneros masculino, feminino e neutro nas terceiras pessoas, o que o tornou bastante artificial e complexo [Pass93]. A separação de gêneros foi estendida também para as palavras sexuadas, que passaram a não ter gênero definido por si mesmas, recebendo-o sempre por meio de um sufixo. Tal comportamento distancia o Ido das línguas naturais, pelo menos as européias. Já o Esperanto usa o modelo pronominal simples do inglês. Na tabela abaixo quando duas opções aparecem separadas por barra então o gênero não tem distinção definida.

Ido	Esperanto	Inglês	Português
il	li	he	ele
el	ŝi	she	ela
lu	li / ŝi	he / she	ele / ela
ol	ĝi	it	(neutro singular)
{ ili eli li oli	ili	they	{ eles elas eles / elas (neutro plural)
patrulo	patro	father	pai
patrino	patrino	mother	mãe
patro	patro / patrino	father / mother	pai / mãe

Outra reforma que complicou ao invés de simplificar a língua foi a omissão da desinência de caso (*n*) que aparece no objeto direto em Esperanto. A marca continua obrigatória, entretanto, se o objeto estiver anteposto ao verbo. Além de tornar a língua mais complicada, esta regra criou uma porta aberta para a construção de frases ambíguas, ainda mais considerando-se que a concordância de número entre substantivo e adjetivo foi retirada. Eis um exemplo citado em [Blan89] que mostra —apesar de à primeira vista o Ido parecer bastante similar ao Esperanto— uma exceção com propósito naturalístico, o plural *siori*, e algumas palavras com grafia diferente (*jurnalo*, *di*, *me*, *lektis*, *serchas*).

“Altestimata siori! En la *jurnalo* di via urbo me lektis, ke vi serchas kontoristo.” (Ido)

“Altestimataj sinjoroj! En la *ĵurnalo* de via urbo mi legis, ke vi serĉas kontoriston.” (Esperanto)

“Estimadíssimos senhores! No jornal de sua cidade eu li que vocês procuram trabalhadores de escritório.”

3.2 Comparação com Línguas Naturais

O Esperanto, na sua condição de língua planejada voltada para a comunicação internacional, tem como uma de suas premissas ser de tal maneira que o aprendizado e o uso sejam mais fáceis do que os de uma língua natural no papel

de língua estrangeira. Isto é realizado através de uma gramática que privilegia a regularidade (característica *a priori*) e de um léxico baseado nas línguas naturais, majoritariamente nas latinas (característica *a posteriori*).

A regularidade da língua implica um menor esforço para aprendê-la e memorizá-la, além de dar mais confiança e segurança ao falante para não cometer erros, dessa maneira evitando situações embaraçosas ou ridículas. Por outro lado a regularidade confere mais precisão e concisão às palavras e ao mesmo tempo mais simplicidade nas derivações. Outra vantagem é que a expansão do léxico é feita através de simples raízes, e não de palavras completas e já derivadas, como nas línguas naturais.

O léxico esperantista é composto principalmente de raízes latinas (cerca de 70% [Blan89]); raízes germânicas e eslavas também compõem uma parte significativa. A presença de grande quantidade de palavras cognatas (pelo menos para falantes de línguas européias) abrevia o tempo de aprendizado. Mesmo falantes de línguas não-européias têm um tempo de aprendizado relativamente curto, devido à regularidade e à presença de muitas palavras “internacionais” (semelhantes em muitas línguas diferentes). Tais palavras puderam ser incorporadas ao Esperanto facilmente, através da criação de uma raiz (com possível adaptação para o alfabeto esperantista), que depois percorre todo o sistema de derivações comum a todas as raízes.

As comparações mostradas abaixo percorrem os três níveis lingüísticos básicos: fonético (Seção 3.2.1), morfológico (Seção 3.2.2) e sintático (Seção 3.2.3). A maior parte é feita entre o Esperanto e o português ou o inglês; às vezes outras línguas também são incluídas nas comparações.

3.2.1 Ortografia

A correspondência entre os sons e as letras usadas para representá-los é perfeita em Esperanto: existem 28 sons e 28 letras (o alfabeto completo aparece no Apêndice B). Em comparação, o português tem 33 sons e 23 letras. O inglês e o francês usam 26 letras, e tem 35 e 36 sons, respectivamente [Pass93]. Como a escrita torna-se uma tarefa que envolve uma dose de memorização significativa, o aprendizado de uma língua estrangeira torna-se penoso, uma vez que casos especiais diversos devem ser percebidos. Alguns exemplos da variação ortográfica presente em línguas naturais demonstram esta dificuldade:

- [s] em português: sapo, massa, cinema, rescindir, moça, exceção, extenso, feliz;
- [z] em português: zebra, casa, exército;
- [i] em inglês: believe, receive, machine, read, see, key, people, we, amoeba;
- [k] em inglês: comic, equal, account, acquire, chaos, black;
- [f] em inglês: sugar, nation, machine, shell;
- [ɛ] em francês: thème, pêche, peigne, jersey, Noël, chaise, maître, Gournay, cachet.

Em comparação, uma pessoa quando aprende Esperanto consegue ler e escrever corretamente quaisquer palavras a partir do momento em que domina corretamente o alfabeto. A sílaba tônica de todas as palavras com pelo menos duas sílabas é sempre a penúltima.

3.2.2 Formação de Palavras

A construção e derivação de palavras em línguas naturais é sujeita a irregularidades em diferentes aspectos, dificultando seu aprendizado como língua estrangeira e seu tratamento computacional. O Esperanto possui uma gramática regular [Wenn01, Matt87] que evidencia-se como fácil de ser utilizada tanto verbal quanto computacionalmente. Nos exemplos de irregularidades abaixo figuram as traduções em Esperanto (em itálico) para servir de comparação.

Formas dos morfemas: um mesmo morfema pode ter variantes de acordo com a palavra na qual aparece ou pode ter alguma parte removida durante a formação de palavras. Estes fenômenos recebem os nomes de *alomorfia* e *truncamento*, respectivamente. Exemplos de alomorfia em português: nadar (*naĝi*), natatório (*naĝa*); perceber (*percepti*), percepção (*percepto*); limão (*citrono*), limoeiro (*citronujo*); incapaz (*nekapabla*), ilegal (*neleĝa*), impróprio (*netaŭga*).

Regras de derivação: pode haver várias regras que usam afixos diferentes, mas que contêm a mesma idéia. Como exemplo eis palavras que significam qualidades, formadas com sufixos diferentes: alegria (*ĝojeco*), solidão (*soleco*), liberdade (*libereco*), nudez (*nudeco*), juventude (*juneco*), doçura (*dolĉeco*). Estas regras de derivação são condicionais, associando os morfemas somente quando eles têm determinadas formas ou características. Assim, em português, redondeza (*rondeco*) e legível (*legebla*) são palavras corretas, mas **quadradeza* (*kvadratoeco*) e **escrevível* (*skribebla*) não são. Da mesma forma em inglês temos bigger (*pli granda*) e biggest (*plej granda*) mas não **goodder* (*pli bona*) e **gooddest* (*plej bona*).

Marcas nominiais de gênero, número e caso: costumam ter regras gerais com um grande número de exceções. Por exemplo, a marca de plural em português e em inglês geralmente é um simples esse, mas existem diversas exceções: mares (*maroj*), anéis (*ringoj*), homens (*homoj*), leões (*leonoj*), pães (*panoj*), porcos⁴ (*porkoj*); children (*infanoj*), feet (*piedoj*), leaves (*folioj*), mice (*musoj*), tomatoes (*tomatoj*).

Conjugações verbais: verbos irregulares são muito comuns nas línguas naturais. O inglês, apesar de ter poucas conjugações, é cheio de irregularidades, como os verbos conjugados no passado a seguir demonstram:⁵ broke (quebrou, *rompis*), sent (enviou, *sendis*), fought (lutou, *luktis*), shook (sacudiu, *skuis*), gave (deu, *donis*), ate (comeu, *manĝis*), saw (viu, *vidis*), slept (dormiu, *dormis*).

Numerais: os numerais cardinais entre 10 e 100 são formados através de regras diversas, que se aplicam a apenas parte dos números da faixa, muitas vezes recorrendo à alomorfa. Por exemplo, em português existe setenta (*sepdek*) mas não *quatrenta (*kvarddek*) e vinte (*dudek*) não tem nenhuma relação com dois (*du*). Uma língua que apresenta muitas arbitrariedades é o francês, no qual os números de 61 a 79 e de 81 a 99 são formados por 60 (ou 80) seguido de 1 a 19: soixante-quatre (sessenta e quatro, *sesdek kvar*), soixante-dix (setenta, *sepdek*), soixante-treize (setenta e três, *sepdek tri*); quatre-vingts (oitenta, *okdek*), quatre-vingt-neuf (oitenta e nove, *okdek naŭ*), quatre-vingt-onze (noventa e um, *naŭdek unu*). Além disso, numerais não-cardinais às vezes têm outras raízes —em português não há relação lexical entre um (*unu*) e primeiro (*unua*), ou entre dois (*du*) e segundo (*dua*).

O processo como os morfemas se juntam varia muito entre as línguas naturais. Basicamente existe uma divisão entre línguas analíticas e sintéticas [Bybe85]. Nas línguas analíticas as palavras primitivas são morfemas, e as palavras derivadas são simples associações de outras palavras, que não sofrem modificações durante a derivação. Exemplos deste tipo de língua são o chinês e o vietnamita. Alguns exemplos de palavras derivadas chinesas (no dialeto mandarim) aparecem abaixo, com a transcrição fonética costumeira desta língua, na qual os tons das vogais são indicados por meio de acentos.

界 [jiè] (limite) + 面 [miàn] (superfície)
 → 界面 [jiè miàn] (interface)

社会 [shè huì] (sociedade) + 学 [xué] (estudo)
 → 社会学 [shè huì xué] (sociologia)

电子 [diàn zǐ] (eletrônico) + 计算 [jì suàn] (calcular) + 机 [jī] (máquina)
 → 电子计算机 [diàn zǐ jì suàn jī] (computador)

Nas línguas sintéticas os morfemas sofrem modificações quando se relacionam uns com os outros na formação das palavras. A maneira como os morfemas são dispostos divide esta categoria em línguas aglutinativas e línguas fusionais.⁶ As línguas aglutinativas constroem seqüências de morfemas, que são afetados por alomorfa e truncamento. Nestas línguas existem morfemas chamados afixos (divididos em prefixos e sufixos, de acordo com sua posição na palavra) que só existem em palavras derivadas, ou seja, que embutem em si regras de derivação. Exemplos de línguas aglutinativas são o português, o inglês e as línguas europeias de modo geral. As línguas fusionais constroem palavras intercalando os morfemas uns nos outros de acordo com esquemas definidos na gramática.

⁴Neste exemplo a pronúncia muda do singular (ò) para o plural (ó).

⁵As traduções em português apresentam a forma da terceira pessoa do singular.

⁶Em [Spro92] é usada uma nomenclatura diferente para línguas analíticas e fusionais: elas são chamadas de línguas *isolativas* e *inflectionais*, respectivamente. Além disso um terceiro tipo de língua sintética é definido (as línguas *polissintéticas*), no qual elementos que costumam ser palavras soltas, como argumentos de verbos, tornam-se morfemas.

Dois exemplos são o árabe e o hebraico. Abaixo estão exemplos de palavras árabes no singular e no plural, citados em [Spro92]. As raízes das palavras estão sublinhadas, destacando a alomorfa que acontece em português.

[jundub] (gafanhoto) → [janaadib] (gafanhotos)

[muutur] (motor) → [mawaatir] (motores)

[sultaan] (sultão) → [salaatiin] (sultões)

Sob o ponto-de-vista “clássico” mostrado em cartilhas, o Esperanto é ensinado como uma língua aglutinativa, com prefixos e sufixos. No entanto estes afixos podem formar palavras por si mesmos, contrariando a definição formal de afixo, e os morfemas são invariáveis, não sofrendo alomorfa nem truncamento. Estas características lingüísticas forçam uma reclassificação do Esperanto como língua *analítica*, embora tenha um léxico proveniente de línguas aglutinativas [Piro01].

3.2.3 Construção de Orações

Também existem idiosincrasias nas línguas naturais na construção de orações, que tornam seu tratamento computacional mais complexo e que também podem dificultar seu aprendizado como língua estrangeira. Em primeiro lugar, uma mesma palavra pode ter classes gramaticais diferentes dependendo da oração. Este fenômeno é comum em português e muito freqüente no inglês, como ressaltado nos exemplos abaixo. A tradução em Esperanto (em itálico) mostra que nesta língua a classe gramatical de cada palavra é marcada por meio de uma desinência específica.⁷

O **policia** ajudou a vítima de roubo.
(*La **policisto** helpis la viktimon de ŝtelo.*)

A violência **policia** foi criticada pelo prefeito.
(*La **policia** perforto estis kritikata de la urbestro.*)

Light is the fastest thing in the universe.
(A **luz** é a coisa mais rápida (dentre todas) no universo.)
(***Lumo** estas la plej rapida afero en la universo.*)

Kangaroos are born extremely small and **light**.
(Cangurus nascem extremamente pequenos e **leves**.)
(*Kanguruoj naskiĝas ekstreme malgrandaj kaj **malpezaj**.*)

Torches **light** our way through the dark forest.
(Tochas **iluminam** nosso caminho através da floresta escura.)
(*Torĉoj **lumas** nian vojon tra la malhela arbaro.*)

⁷As desinências verbais são mostradas na Seção 3.1.4. As outras desinências são: **o** para substantivos, **a** para adjetivos e **e** para advérbios. Existem determinadas raízes “primitivas” que podem aparecer sem desinência, como pronomes pessoais, preposições, numerais e conjunções.

Interrogações em línguas (como o português) que não têm uma palavra marcadora de interrogação, nem dispõem da possibilidade de inverter a ordem das palavras, dependem totalmente da entoação correta. No entanto ao se falar uma língua estrangeira nem sempre se consegue a entoação adequada, o que pode causar desentendimentos. Em virtude disto o Esperanto possui a conjunção *ĉu* para marcar a interrogação. Como visto nos exemplos abaixo [Pass93], esta palavra também é usada nas interrogações indiretas, enquanto que nas cinco línguas naturais mostradas é usada a conjunção que representa condição.

Mi demandis lin **ĉu** Petro venos morgaŭ. (Esperanto)
 Io gli ho domandato **se** Pietro verrà domani. (Italiano)
 Yo le he preguntado **se** Pedro vendrá mañana. (Espanhol)
 Je lui ai demandé **se** Pierre viendrait demain. (Francês)
 I asked him **if** Peter would come tomorrow. (Inglês)
 Eu perguntei-lhe **se** Pedro virá amanhã.

Construções sintáticas ambíguas aparecem freqüentemente nas línguas naturais. O Esperanto também possui ambigüidades sintáticas, mas elas são menos freqüentes, devido à presença da marca de caso **n** (cujo uso principal é destacar o objeto direto das orações) e de construções sintáticas claras. As orações abaixo, por exemplo, não são ambíguas em Esperanto:

Eu o elegi como presidente. = $\left\{ \begin{array}{l} \text{Eu o elegi como presidente (que sou).} \\ \text{(Mi elektis lin kiel prezidanto.)} \\ \\ \text{Eu o elegi como presidente (que ele é).} \\ \text{(Mi elektis lin kiel prezidanton.)} \end{array} \right.$

He has been a japanese war camp prisoner.
 = $\left\{ \begin{array}{l} \text{He has been a prisoner of a japanese war camp.} \\ \text{(Ele foi um prisioneiro de um campo militar japonês.)} \\ \text{Li estis malliberulo de japana milita kampo.} \\ \\ \text{He has been a camp prisoner during the japanese war.} \\ \text{(Ele foi prisioneiro de campo durante a guerra japonesa.)} \\ \text{Li estis kampa malliberulo dum la japana milito.} \\ \text{Li estis kampa malliberulo japanmilito.} \end{array} \right.$

Uma outra diferença entre o Esperanto e as línguas naturais diz respeito à liberdade de disposição das palavras dentro dos sintagmas. Como as palavras em Esperanto mostram explicitamente atributos sintáticos —através das designações de classe gramatical e de caso— sua ordem pode ser muito menos rígida do que a de uma língua natural, característica esta utilizada amplamente na poesia. Uma ordem usual é seguida normalmente na prosa, mas o desvio desta ordem possibilita uma grande liberdade para expressar ênfase nos vários elementos. Outra vantagem é que pessoas que estão começando a aprender Esperanto

podem usar a ordem costumeira de sua língua materna sem correr o risco de não serem compreendidos, ainda que falem de uma maneira não costumeira.

4 Resumo Gramatical do Esperanto

Um resumo abrangente da gramática do Esperanto não pode ser feito no espaço disponível neste artigo,⁸ e acabaria mudando seu foco de uma abordagem mais computacional para uma mais lingüística. Sendo assim, a seguir é feito um apanhado geral (baseado em [Matt87]) nos níveis morfológico (Seção 4.1) e sintático (Seção 4.2) somente para fixar as explicações já mostradas anteriormente e para acrescentar novas que são exploradas na Seção 5. O nível fonético já é devidamente tratado em outros lugares: os sons utilizados no Esperanto estão exemplificados no Apêndice B, e alguns comentários sobre sua fonética e sua ortografia são apresentados na Seção 3.2.1.

4.1 Morfologia

As palavras simples em Esperanto são compostas pela justaposição de uma *raiz* e de uma *desinência*. A raiz contribui com o sentido —uma coisa, uma ação, uma qualidade etc.— enquanto a desinência marca a classe gramatical. A tabela abaixo mostra palavras formadas a partir da raiz *telefon*:

Palavra	Análise	Tradução
telefono	telefon + o	telefone
telefona	telefon + a	telefônico
telefone	telefon + e	telefonicamente
telefoni	telefon + i	telefonar

Algumas raízes são chamadas de *afixos* (divididos em prefixos e sufixos, dependendo de sua posição), pois aparecem quase sempre somente na derivação de palavras a partir de outras raízes. O Esperanto possui muitos sufixos e alguns prefixos, cada qual com seu significado particular coerente e fixo, que permitem que uma grande quantidade de palavras possa ser derivada a partir de uma única raiz, a partir de aplicações sucessivas de afixos. Eis alguns exemplos de derivações:

Palavra	Análise	Tradução
amikino	amik + in + o (amigo + feminino + subst.)	amiga
stelaro	stel + ar + o (estrela + conjunto + subst.)	constelação
preĝejo	preĝ + ej + o (rezar + lugar + subst.)	igreja
maldekstra	mal + dekstr + a (contrário + direito + adj.)	esquerdo
disdoni	dis + don + i (dispersão + dar + verbo)	distribuir

Além do processo de derivação também existe o de composição, no qual duas raízes se juntam para formar uma palavra composta. A raiz que fica à

⁸Uma tentativa de fazer um resumo mais abrangente encontra-se em [Ribe01].

direita na composição é o elemento principal. A raiz à esquerda serve para de alguma forma tornar preciso o sentido geral do elemento principal. A seguir são mostrados exemplos de composições:

Palavra	Análise	Tradução
skribtablo	skrib + tabl + o (escrever + mesa + subst.)	escrivadinha
vaporŝipo	vapor + ŝip + o (vapor + navio + subst.)	navio a vapor
dikfingro	dik + fingr + o (grosso + dedo + subst.)	polegar
novluno	nov + lun + o (novo + lua + subst.)	lua nova

Outro processo morfológico presente no Esperanto subordina um sintagma de poucas palavras num morfema, acrescentando em seguida a desinência de classe gramatical correspondente. Tal processo permite que o texto fique mais ou menos sintético, de acordo com o estilo do autor. Abaixo estão alguns exemplos.

Palavra	Análise	Tradução
tiumaniere	[tiu maniero] + e ([essa maneira] + adv.)	nessa maneira
libertempi	[libera tempo] + i ([tempo livre] + verbo)	tirar férias
senutila	[sen utilo] + a ([sem utilidade] + adj.)	inútil
internacia	[inter nacioj] + a ([entre nações] + adj.)	internacional

4.2 Sintaxe

A oração em Esperanto normalmente usa a ordem sujeito-verbo-objeto. No entanto esta ordem não é obrigatória, já que o objeto direto é marcado com a desinência de caso **n** nos substantivos e adjetivos. Isto faz com que a língua admita variações na ordem das palavras, com o intuito de enfatizar uma ou outra.

Para formar a oração negativa a partir da afirmativa basta acrescentar a palavra *ne* (não) antes do verbo. As orações também se tornam negativas quando contêm certas palavras inerentemente negativas, não podendo apresentar dupla negação. Eis exemplos de orações negativas:

Oração	Tradução
Ne veku la infanon.	Não acorde a criança.
Ŝi ne forgesis mian datrevenon.	Ela não esqueceu meu aniversário.
Mi neniel komprenis tion.	Eu não entendi isso de jeito nenhum .
Mi vokigis neniun .	Não mandei chamar ninguém .

As orações interrogativas que demandam uma resposta “sim” ou “não” recebem a palavra *ĉu* no começo. Interrogações que demandam respostas específicas usam as palavras interrogativas adequadas à pergunta, como mostrado nos exemplos abaixo.

Oração	Tradução
Ĉu pluvas?	Está chovendo?
Ĉu vi ŝatas glaciaĵon?	Você gosta de sorvete?
Kiu malfermis la fenestron?	Quem abriu a janela?
Kiam vi alvenos?	Quando você vai chegar?

A voz passiva é feita trazendo-se o objeto para a posição de sujeito e mudando-se o verbo para *esti* (estar) seguido de um particípio. O agente da passiva é marcado com a preposição *de*. Dois pares de orações correspondentes são mostrados a seguir.

Oração ativa	Oração passiva
La knabino vidas la domon. (A menina vê a casa.)	La domo estas vidata de la knabino. (A casa é vista pela menina.)
Hundo vundis la katon. (Um cachorro feriu o gato.)	La kato estis vundata de hundo. (O gato foi ferido por um cachorro.)

4.2.1 Papéis Sintáticos dos Sintagmas

As partes que podem compor uma oração têm diferentes papéis sintáticos em relação ao verbo. Estes papéis são indicados através da presença ou da ausência da marca de caso **n** e de preposições. Existem cinco papéis sintáticos: sujeito, objeto, predicativo, vocativo e complemento.

Em três papéis sintáticos (sujeito, predicativo e vocativo) a marca de caso sempre está ausente. O sujeito faz a ação indicada pelo verbo. O predicativo mostra (por meio de um verbo) uma qualidade do sujeito ou do objeto. O vocativo identifica a pessoa com quem se fala. Abaixo se encontram pares de exemplos destes três papéis, na ordem em que foram apresentados.

Oração	Tradução
La birdoj flugis.	Os pássaros voaram.
Mi vojaĝos al Parizo.	Eu vou viajar para Paris.
Ĉiuj boteloj estis malplenaj .	Todas as garrafas estavam vazias .
Mi trovis la vinon bona .	Eu achei bom o vinho.
Maria , bonvole venu ĉi tie.	Maria , por favor venha aqui.
Ĉu vi volas akvon, filo ?	Você quer água, filho ?

O objeto (direto) recebe a ação do verbo, e é marcado com a desinência de caso nos substantivos e adjetivos, como mostrado nos exemplos abaixo.

Oração	Tradução
Hodiaŭ mi tagmanĝis rizon .	Hoje eu almocei arroz .
Johano aĉetis mian domon .	João comprou minha casa .
La kato kaptis la muson .	O gato pegou o rato .

O complemento mostra as circunstâncias da ação do verbo. Pode haver complementos de tempo, de lugar, de modo etc., que são realizados por meio de

sintagmas adverbiais ou preposicionados, como visto a seguir.

Oração	Tradução
Mi vizitos vin en marto .	Eu vou te visitar em março .
Mi renkontis ŝin en la klubo .	Eu encontrei com ela no clube .
Li subite aperas.	Ele apareceu subitamente .

Orações subordinadas podem perfazer papéis sintáticos diversos. Em Esperanto é costume separar com vírgulas as orações subordinadas.

Oração	Tradução	Papel
Ŝajnas, ke pluvos .	Parece que vai chover .	sujeito
Li diris, ke li estas malsata .	Ele disse que está com fome .	objeto
Kiam li venis , ŝi foriris.	Quando ele veio , ela partiu.	compl.

4.2.2 Observações sobre a Construção de Sintagmas

Um sintagma é composto de um núcleo (palavra principal) e de adjuntos, que têm algum tipo de relação com o núcleo. A classificação do sintagma é feita de acordo com a categoria gramatical de seu núcleo. Os adjuntos são normalmente opcionais e constituem sintagmas por si mesmos. Abaixo se encontram algumas observações importantes, para os falantes de português, sobre a construção dos sintagmas em Esperanto.

Sintagma nominal Num sintagma nominal o(s) adjetivo(s) concorda(m) em número e caso com o substantivo. Existe uma aparente exceção a esta regra, que acontece quando o adjetivo é um predicativo de objeto. Neste caso o substantivo recebe a marca de caso, por ser objeto, mas o adjetivo não, por ser predicativo. Eis alguns exemplos contrastantes, com o sintagma nominal no papel de objeto ressaltado:

Oração	Tradução
Mi trovis la libron tedema.	Eu encontrei o livro (que acho) tedioso.
Mi trovis la libron tedeman .	Eu encontrei o livro (que é) tedioso .
Mi pentris la muron blanka.	Eu pintei a parede de branco.
Mi pentris la muron blankan .	Eu pintei a parede (que era) branca .

Não existe artigo indefinido em Esperanto. A ausência do artigo definido *la* é equivalente ao artigo indefinido, se não existir nenhuma outra palavra que defina o substantivo, como exemplificado abaixo.

Oração	Tradução
Knabo falis.	Um menino caiu.
La knabo kuras.	O menino corre.
Infano foriris.	Uma criança foi embora.
Sinjoro Petro foriris.	O senhor Pedro foi embora.

Sintagma adjetivo Seguindo mais uma vez princípios de regularidade, o adjetivo em Esperanto nunca tem a função de advérbio. Abaixo o contraste entre adjetivo e advérbio é ressaltado por meio de exemplos.

Oração	Tradução
Aviadiloj estas rapidaj .	Aviões são rápidos .
Aviadiloj flugas rapide .	Aviões voam rápido (= rapidamente).
Parolu per laŭta voĉo.	Fale em voz alta .
Parolu laŭte .	Fale alto (= de modo alto).

Sintagma adverbial O sintagma adverbial pode caracterizar tempo, lugar, modo, causa, intensidade etc. O advérbio de lugar pode receber a desinência de caso, que neste caso indica *direção*, como ressaltado nos exemplos abaixo.

Oração	Tradução
Li paŝas hejme .	Ele caminha em (dentro de) casa .
Li paŝas hejmen .	Ele caminha para casa .
Eskimoj vivas norde .	Esquimós vivem no norte .
Ni iru norden .	Vamos para o norte .

Sintagma preposicionado As preposições do Esperanto têm sentidos bem definidos e fixos.⁹ Se o sintagma preposicionado indicar lugar seu complemento pode receber a marca de caso, de modo análogo ao que ocorre no sintagma adverbial, como visto a seguir.

Oração	Tradução
La kato kuras en la ĉambro .	O gato corre no (dentro do) quarto .
La kato kuras en la ĉambrom .	O gato corre para o quarto .
Ne parku antaŭ la garaĝo .	Não estacione na frente da garagem .
Ni paŝis antaŭ la domon .	Nós andamos para a frente da casa .

5 Tratamento Computacional do Esperanto

O processamento lingüístico computacional aplicado às línguas naturais também pode ser aplicado a textos em Esperanto. Tal processamento envolve análises gramaticais bastante diferentes (Seção 5.1) nos níveis fonético, morfológico e sintático. Algumas sugestões de aplicações para as análises lingüísticas são apresentadas na Seção 5.2.

⁹Existe uma preposição (*je*) que é usada nas situações em que nenhuma das outras passaria o sentido correto.

5.1 Análises Lingüísticas Básicas

O processamento lingüístico computacional do Esperanto é mais fácil, em relação às línguas naturais, pois tem menos casos para serem tratados, por causa das regularidades apresentadas nas Seções 3.2 e 4. Além disso o dicionário pode ser bastante reduzido em comparação com dicionários utilizados no processamento de línguas naturais, por causa da ausência de alomorfia e truncamento. Um dicionário de morfemas já basta para realizar a análise morfológica.

A seguir são propostos procedimentos para realizar as análises fonética (Seção 5.1.1), morfológica (Seção 5.1.2) e sintática (Seção 5.1.3) em Esperanto. As inter-relações entre as análises são apontadas na Seção 5.1.4.

5.1.1 Análise Fonética

A análise fonética trata da correspondência entre os sons (*fonemas*) e os símbolos gráficos (*grafemas*) de uma língua. Seus três objetivos principais são: descobrir a seqüência de sons das palavras a partir de suas letras, a maneira como estes sons são agrupados em sílabas, e qual sílaba de cada palavra recebe ênfase na pronúncia (a sílaba *tônica*). Estes aspectos são os normalmente ressaltados na transcrição fonética mostrada em dicionários.

Esta análise aplicada ao Esperanto se revela quase trivial, pelo menos para estes objetivos principais, uma vez que em Esperanto o alfabeto é fonético e a sílaba tônica é fixa (sempre a penúltima em palavras com mais de uma sílaba). O número de sílabas corresponde ao número de vogais numa palavra. A separação das palavras em sílabas pode ser realizada mediante a definição de regras de associação das consoantes às vogais e às próprias consoantes entre si (encontros consonantais). Descobrir figuras de estilo —por exemplo se uma oração é uma aliteração, como *o rato roeu a roupa do rei*— envolve uma análise mais aprofundada e com maior dificuldade.

Eis a análise fonética de *Gratulon al vi* (Parabéns a você),¹⁰ de acordo com a transcrição fonética mostrada no Apêndice B. As sílabas são separadas por pontos, e a sílaba tônica é precedida de um apóstrofo.

En la plej bela tago	[en la plej 'be.la 'ta.go]	(No dia mais bonito
ni gratulas al vi,	[ni gra.'tu.las al vi]	nós te parabenizamos,
dezirante feliĉon	[de.zi.'ran.te fe.'li.tʃon]	desejando felicidade
kaj la benon de Di'!	[kaj la 'be.non de di]	e a bênção de Deus!)

5.1.2 Análise Morfológica

Embora a palavra seja a unidade mínima de comunicação, ela ainda pode ser subdividida em partes menores, denominadas *morfemas*. A maneira como os

¹⁰O apóstrofo pode ser usado em Esperanto para remover a desinência *o* dos substantivos e o *a* final do artigo definido *la*. Isto ocorre somente na poesia. A vogal tônica permanece a mesma de antes.

morfemas se relacionam lembra o modo como as palavras se juntam para formar um sintagma: também existe um elemento central que caracteriza todo o conjunto e que se liga aos outros elementos, que agem como se fossem adjuntos. No entanto o processo morfológico é fundamentalmente diferente do sintático, pois envolve regras que não têm paralelo na sintaxe —por exemplo, um sufixo pode se ligar somente a raízes com uma certa forma, mas não com outras, como mostrado na Seção 3.2.2. A análise morfológica consiste na determinação dos morfemas que constituem uma palavra e de suas relações. Seu resultado, portanto, é uma estrutura hierárquica que relaciona os morfemas. Abordagens para realizar análise morfológica por computador encontram-se em [Spro92].

Dividir uma palavra em Esperanto nos possíveis morfemas que a constituem é uma tarefa relativamente simples: basta partir a palavra em pedaços e verificar se cada um é um morfema válido. Normalmente este processo pode levar a várias possibilidades, cuja análise de acordo com as regras morfológicas [Wenn01, Matt87] leva até a alternativa correta. Nos casos em que realmente ocorre homografia a análise pode levar a alternativas diferentes. Embora quase sempre apenas uma seja usada na prática, todas devem ser consideradas válidas em princípio. Eis alguns exemplos:

$$\begin{aligned}
 \text{jaro} &= \begin{cases} \text{jar/o (ano/substantivo)} \Rightarrow \text{correta} \\ \text{j/ar/o (plural/conjunto/substantivo)} \Rightarrow \text{inválida} \end{cases} \\
 \text{paroli} &= \begin{cases} \text{parol/i (falar/verbo)} \Rightarrow \text{correta} \\ \text{par/ol/i (par/do que/verbo)} \Rightarrow \text{inválida} \\ \text{par/o/li (par/substantivo/ele)} \Rightarrow \text{inválida} \end{cases} \\
 \text{lasi} &= \begin{cases} \text{las/i (deixar/verbo)} \Rightarrow \text{correta} \\ \text{la/si (artigo/pronome reflexivo)} \Rightarrow \text{inválida} \end{cases} \\
 \text{mangarbo} &= \begin{cases} \text{mang/arb/o (manga/árvore/substantivo)} \Rightarrow \text{correta} \\ \text{man/garb/o (mão/feixe/substantivo)} \Rightarrow \text{improvável} \end{cases}
 \end{aligned}$$

As relações entre os morfemas podem ser determinadas de acordo com suas propriedades. Primeiramente a divisão básica dos morfemas em raízes, afixos e desinências já evidencia tipos diferentes de relações. A classe gramatical *intrínseca* de cada raiz (reconhecida primeiramente em [Saus16]) tem grande importância nos processos de formação de palavras, notadamente no de composição [Schu89b]. As raízes por si mesmas podem ser substantivas (como *libr*, *tabl*, *krajon* —respectivamente livro, mesa e lápis), adjetivas (*bel*, *alt*, *util* —bonito, alto, útil) ou verbais (*manĝ*, *rid*, *mensog* —comer, rir, mentir). Assim é preciso distinguir entre a categoria gramatical da palavra como um todo, determinada pela desinência, e a categoria gramatical da própria raiz. [Saus16] usa o termo *gramatika karaktero* (caráter gramatical) para denominar esta última.

As relações são sempre entre apenas dois morfemas de cada vez, que depois agem como um conjunto em relações subseqüentes. Dessa maneira a estrutura hierárquica que resulta da análise morfológica do Esperanto é uma simples árvore binária. Abaixo aparecem alguns exemplos de análise, nos quais os parênteses representam um nó da árvore que leva aos dois elementos dentro dos parênteses.

Palavra	Análise	Tradução
dormi	(dorm + i)	dormir
laboro	(labor + o)	trabalho
ekvidi	((ek + vid) + i)	avistar
arbaro	((arb + ar) + o)	floresta
internacia	((inter + naci) + a)	internacional
altmontaro	((alt + mont) + ar) + o)	cordilheira
sekvinbero	((sek + (vin + ber)) + o)	uva-passa

A análise de uma palavra pode aproveitar análises de palavras mais simples, feitas anteriormente, para decidir por uma alternativa. Seja as seguintes palavras abaixo:

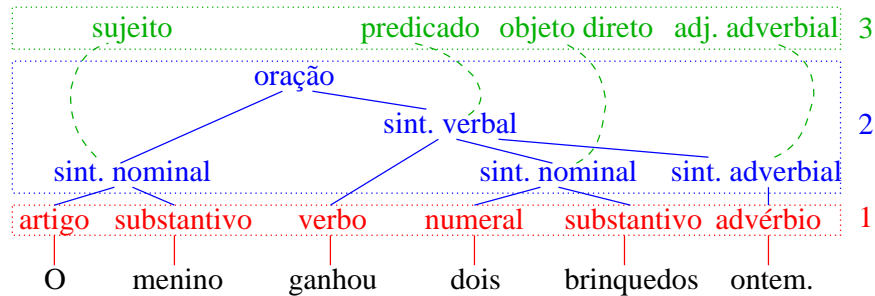
$$\begin{aligned} \text{fervojisto (trabalhador de ferrovia)} &= \begin{cases} ((\text{fer} + (\text{voj} + \text{ist})) + \text{o}) \\ (((\text{fer} + \text{voj}) + \text{ist}) + \text{o}) \end{cases} \\ \text{fervojestacidomo (estação ferroviária)} &= \begin{cases} ((((\text{fer} + \text{voj}) + \text{staci}) + \text{dom}) + \text{o}) \\ ((\text{fer} + (\text{voj} + (\text{staci} + \text{dom}))) + \text{o}) \\ (((\text{fer} + (\text{voj} + \text{staci})) + \text{dom}) + \text{o}) \\ ((\text{fer} + ((\text{voj} + \text{staci}) + \text{dom})) + \text{o}) \\ (((\text{fer} + \text{voj}) + (\text{staci} + \text{dom})) + \text{o}) \end{cases} \end{aligned}$$

Como existem as palavras *fervojo* (ferrovia) e *stacidomo* (estação), as análises das palavras acima podem escolher a segunda e a quinta opções, respectivamente. Casos como *fervojestacidomo*, nos quais várias raízes são compostas sucessivamente, não são muito comuns em Esperanto. Normalmente acontecem apenas derivações sucessivas, isto é, palavras com vários afixos.

5.1.3 Análise Sintática

O processo de determinar a estrutura hierárquica subjacente à seqüência linear das palavras (identificando a categoria gramatical de cada uma e suas relações) é chamado de análise sintática. O léxico de uma língua define a qual(is) categoria(s) uma palavra pode pertencer, enquanto a sintaxe restringe as relações possíveis entre as categorias. Tais relações se mostram nas palavras através de concordância, ordem linear e marcas de caso, por exemplo.

A análise sintática pode ser feita em três níveis diferentes. No nível mais baixo ou *categorial* é reconhecida apenas as categorias gramaticais as quais as palavras pertencem. O nível intermediário ou *sintagmático* agrupa as palavras em sintagmas, que recebem classificações distintas de acordo com a categoria de sua palavra principal. Finalmente o nível mais alto ou *funcional* reconhece qual é o papel sintático dos sintagmas, isto é, se o sintagma funciona como sujeito, objeto etc. A figura abaixo mostra estes três níveis na oração *O menino ganhou dois brinquedos ontem*.



A análise sintática computacional é tradicionalmente feita através de regras que relacionam sintagmas com seus elementos, determinando a posição de cada um. As regras são testadas em cascata a fim de saber se casam com as palavras sob análise. Tal procedimento envolve basicamente tentativa e erro, de forma que a análise procede até o ponto no qual nenhuma regra pode ser aplicada, quando então é feito um retrocesso até um ponto em que é possível aplicar uma regra diferente e seguir em busca de uma possível solução. De acordo com a direção em que as regras são percorridas define-se dois tipos de análise: a análise *ascendente* ou *bottom-up* parte das palavras até chegar ao símbolo inicial da gramática; a análise *descendente* ou *top-down* parte do símbolo inicial até chegar às palavras [Alle87].

Como a análise ascendente procura primeiro encontrar a categoria gramatical de cada palavra para depois relacionar palavras próximas em sintagmas, este tipo de análise parece ser mais adequado para analisar o Esperanto, uma vez que encontrar as categorias gramaticais das palavras —ou seja, realizar a análise no nível categorial— é um processo trivial, já que a categoria é marcada léxica ou morfológicamente. As análises nos níveis mais altos podem então ser melhor direcionadas tendo esta informação como base. Como a posição das palavras em Esperanto não é rígida (existem poucas regras que usam a ordem das palavras) é necessário que as regras de análise sejam “elásticas”, com alguns elementos sem posições claramente definidas.

O processo de análise constrói a *árvore de derivação* que representa a estrutura sintática da oração. O fato de se trabalhar no máximo no nível sintático funcional admite a ocorrência de ambigüidades —embora em menor frequência do que nas línguas naturais, como visto na Seção 3.2.3— que só podem ser resolvidas por meio de informações semânticas, portanto além da capacidade de resolução da análise sintática. Neste caso a análise deve dar como resultado todas as possibilidades.

Já existe um sistema voltado para a análise sintática do Esperanto disponível para testes via internet. Ele faz parte do projeto VISL [Bick00], que engloba um conjunto de ferramentas para análise sintática e tradução automática que trabalha com várias línguas. Não há muitas informações técnicas disponíveis, mas revela-se que o analisador é uma adaptação do utilizado para tratar o português.

5.1.4 Relações entre as Análises

As análises propostas acima tratam dos três níveis lingüísticos básicos (fonético, morfológico e sintático). A análise de um nível pode se inter-relacionar com a(s) do(s) nível(is) adjacente(s).

Influência da análise morfológica na análise sintática A análise morfológica provê informações de suma importância para a análise sintática. Em primeiro lugar, ela identifica a classe gramatical de todas as palavras. As desinências de número e caso servem para marcar a concordância entre substantivos e adjetivos. Os papéis sintáticos da oração podem ser identificados de acordo com a presença ou a ausência da marca de caso. Palavras primitivas como conjunções, preposições e o artigo identificam começos de sintagmas.

Se for adicionada informação lexical verbal, é ainda possível identificar na análise morfológica a regência verbal, isto é, quantos e quais são os argumentos que o verbo exige. Esta característica é intrínseca a cada raiz verbal e pode ser alterada pelos sufixos **ig** (que torna o verbo transitivo) e **iĝ** (que o torna intransitivo), como mostrado nos exemplos a seguir:

sidi (estar sentado) → sidigi (mandar sentar, fazer sentar)

brili (arder) → briligi (fazer arder = queimar)

malfermi (abrir) → malfermiĝi (abrir-se sozinho)

detrui (destruir) → detruiĝi (destruir-se)

Influência da análise sintática na análise morfológica A análise sintática muitas vezes pode validar subordinações de sintagmas em morfemas, como as mostradas na Seção 4.1. Mais informação pode ser necessária, no entanto, já que nestas subordinações só permanecem os elementos essenciais do sintagma [Wenn01].

Influência da análise fonética na análise morfológica Às vezes as desinências de classe são usadas para melhorar a eufonia de termos compostos nos quais a pronúncia seria difícil —por exemplo em *noktomanĝi* = nokt/o/manĝ/i (comer à noite = cear). A desinência não é arbitrária, sendo sempre de acordo com o tipo de composição [Wenn01]. A análise fonética pode revelar quais encontros consonantais são considerados difíceis, informando para a análise morfológica em quais pontos esta construção pode ser aceita.

Influência da análise morfológica na análise fonética A análise fonética pode destacar os morfemas de uma palavra composta atribuindo sílabas *subtônicas* às sílabas que seriam tônicas em palavras simples com estes morfemas. As posições destas sílabas subtônicas podem ser determinadas pela análise morfológica.

5.2 Aplicações das Análises Lingüísticas

A seguir são listadas oito aplicações das análises mostradas na Seção 5.1, numa ordem tal que cada uma pode usar as anteriores como base.

5.2.1 Procura “Inteligente” de Texto

Aproveitando as regularidades gramaticais é possível incrementar a busca de palavras ou orações em textos em Esperanto utilizando as análises morfológica e sintática. É possível procurar automaticamente qualquer substantivo ou adjetivo com ou sem a marca de plural e/ou a de caso, ou ainda um verbo em qualquer uma de suas conjugações. Também é possível fazer buscas a partir de um único morfema, procurando derivações e composições —por exemplo, ao se procurar *tuk* (toalha) pode-se encontrar *nasztuko* (toalha de nariz = lenço), *littuko* (toalha de cama = lençol), *mantuko* (toalha de mão) etc.

Se a busca envolver uma oração em vez de palavras, pode-se procurá-la sem manter a ordem fixa. O oração pode ser encontrada tendo o sujeito no final, por exemplo. A mesma oração pode ser procurada simultaneamente nas formas afirmativa, negativa ou interrogativa, nas vozes ativa e passiva. Em vez de orações podem ser procurados sintagmas, e a busca pode encontrar o sintagma usando uma ordem diferente para as palavras.

5.2.2 Análise Enfática

As orações e sintagmas em Esperanto têm uma ordem costumeira entre seus elementos, que no entanto não é rígida. Esta ordem usual pode ser modificada com a intenção de dar maior ênfase num elemento ou em outro. Assim é possível, a partir da análise sintática, realizar uma análise enfática do texto, comparando a ordem apresentada pelos elementos à ordem comum. Num sintagma nominal, por exemplo, um adjetivo posposto ao substantivo demonstra a ênfase no adjetivo. Numa oração, se o objeto precede o sujeito então a ênfase da oração como um todo está no objeto.

5.2.3 Síntese de Voz

O uso mais óbvio da síntese de voz computadorizada é para leitura de textos para cegos. Outras aplicações são possíveis, como leitura de mensagens eletrônicas em telefones celulares, por exemplo. A síntese de voz a partir de textos em Esperanto tem a vantagem de usar uma análise fonética simples (Seção 5.1.1). Informações vindas da análise morfológica podem influir na pronúncia das palavras, como visto na Seção 5.1.4. A análise enfática vista acima pode ser usada para dar o ritmo e a entonação corretos à voz sintetizada.

5.2.4 Análise Semântica

A análise semântica pode ser feita tendo por base a análise morfológica ou a sintática. A partir da análise morfológica pode-se verificar a relação entre os mor-

femas de uma palavra e tentar descobrir seu significado, principalmente no caso das derivações e subordinações de sintagmas. Como a diversidade de sentido numa raiz é pouco comum (embora obviamente exista a noção de sentido próprio e sentido figurado) e as regras de derivação são claras e regulares, as palavras tendem a ter um sentido mais preciso, e assim o computador pode adivinhar seu significado com razoável certeza. Eis as circumlocuções correspondentes aos exemplos de palavras não-simples mostradas na tabela da Seção 5.1.2:

Palavra / análise	Tradução
ekvidi = komenci vidi	avistar = começar a ver
arbaro = aro da arboj	floresta = conjunto de árvores
internacia = inter nacioj	internacional = entre nações
altmontaro = aro da altaj montoj	cordilheira = conjunto de montes altos
sekinbero = seka vinbero	uva-passa = uva seca

A análise semântica pode ser aplicada ao resultado da análise sintática para determinar os papéis semânticos dos sintagmas em relação ao verbo (agente, objeto, beneficiário, origem, destino, tempo, instrumento etc.). Eis dois exemplos:

$$\begin{array}{l}
 \text{Mi aĉetis panojn por la knabinoj.} \\
 \text{(Eu comprei pães para as meninas.)}
 \end{array}
 \left\{ \begin{array}{l}
 \text{mi = agente} \\
 \text{panojn = objeto} \\
 \text{la knabinoj = beneficiário}
 \end{array} \right.$$

$$\begin{array}{l}
 \text{Ŝi veturis al Lisbono per ŝipo.} \\
 \text{(Ela viajou para Lisboa de navio.)}
 \end{array}
 \left\{ \begin{array}{l}
 \text{ŝi = agente} \\
 \text{Lisbono = destino} \\
 \text{ŝipo = instrumento}
 \end{array} \right.$$

A ordem das palavras em Esperanto não têm influência na sua semântica, ao contrário do que ocorre em determinados casos do português —por exemplo, *homem grande* quer dizer algo diferente de *grande homem* (em Esperanto estes dois sentidos têm palavras diferentes: *granda homo* e *grava homo*, respectivamente).

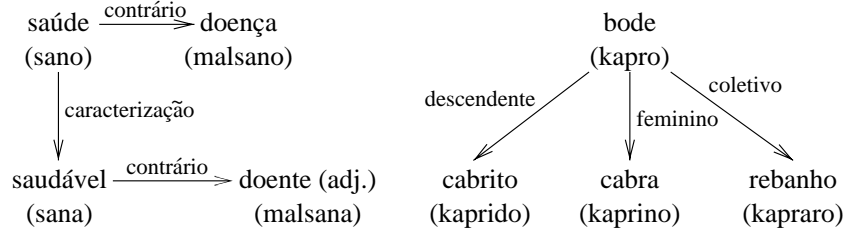
5.2.5 Interfaces Textuais

A análise morfo-sintática do Esperanto abre caminho para a implementação de interfaces com o usuário baseadas em instruções ou diálogos textuais. A aplicação de análise semântica torna o entendimento das instruções mais preciso e menos dependente das palavras exatas usadas pelo usuário: uma entidade poderia ser referida usando o singular, o plural ou um coletivo, por exemplo. Sistemas especialistas, jogos e bancos de dados são exemplos de aplicações que poderiam usar este tipo de interface.

5.2.6 Construção de Dicionários Bilíngües

Relações entre palavras de uma língua natural podem ser construídas usando como base as próprias regras de derivação e inflexão do Esperanto. Estas

relações podem representar regras gerais, exceções às regras gerais ou ainda ligações entre duas palavras não relacionadas lexicamente. Elas formam uma rede semântica —com sinônimos, antônimos, plurais, mudanças de gênero, coletivos etc.— que produz um dicionário bilíngüe Esperanto-língua natural com precisão significativa. A figura abaixo ilustra estas relações entre palavras portuguesas.



Existe um projeto que segue esta linha de raciocínio. O Ergane [Wilg00] é um dicionário multilíngüe voltado para viajantes, que realiza traduções de palavras soltas em dois passos, primeiro da língua de origem para o Esperanto e depois do Esperanto para a língua de destino. As listas bilíngües de palavras vêm separadas do programa, de forma que o usuário só precisa usar o que for necessário. O conteúdo das listas pode ser utilizado livremente.

5.2.7 Resolução de Referências

Um problema sintático importante, que não foi abordado na Seção 5.1.3, é a resolução das referências (pronomes e elipses, principalmente) encontradas no texto. Os dois tipos de referência aparecem nos exemplos abaixo. A elipse funciona como um pronome nulo.

Oração	Tradução
Petro pensa, ke Maria ama lin. =	Pedro acha que Maria o ama. =
Petro pensa, ke Maria ama Petron.	Pedro acha que Maria ama Pedro.
Li kuris, stumblis kaj falis. =	Ele correu, tropeçou e caiu. =
Li kuris, li stumblis kaj li falis.	Ele correu, ele tropeçou e ele caiu.

As referências sempre são difíceis de serem resolvidas computacionalmente [Alle87] por envolverem regras fortemente semânticas. A análise enfática pode auxiliar neste trabalho, já que os pronomes costumam se referir ao elemento que recebe ênfase.

5.2.8 Tradução Automática para Línguas Naturais

Uma base para um sistema de tradução automática do Esperanto para uma língua natural consiste num conjunto de analisadores lingüísticos —morfológico (Seção 5.1.2), sintático (Seção 5.1.3), enfático (Seção 5.2.2), semântico (Seção 5.2.4) e referencial (Seção 5.2.7)— aliado a dicionários bilíngües (Seção 5.2.6). A tradução no outro sentido é muito mais difícil, pois envolve a análise de uma

língua natural e suas irregularidades e arbitrariedades. Não obstante, dois projetos de tradução automática entre línguas naturais utilizam versões modificadas do Esperanto como língua-ponte para realizar traduções em dois passos: DLT e UTL.

O DLT (Distributed Language Translation) [Sadl89, Sadl91, Schu96] foi um projeto de tradução automática multilíngue realizado na Holanda no período de 1984 a 1990, pela empresa Buro voor Systeemontwikkeling. Parte do financiamento do projeto veio do governo holandês. Os dois passos da tradução são independentes e podem ser realizados por diferentes computadores em rede. As razões para as modificações no Esperanto são duas, ambas envolvendo a resolução de ambigüidades:

- para evitar a ocorrência de homografia, os morfemas são claramente separados na escrita;
- para evitar ambigüidades sintáticas, algumas palavras novas são introduzidas, e algumas regras sintáticas são alteradas.

Um protótipo foi completado em 1988 para traduzir entre inglês e francês, usando um banco de dados para tratar ambigüidades. O banco contém duplas de palavras relacionadas, que auxiliam na escolha de uma entre várias traduções possíveis. Um algoritmo foi desenvolvido para procurar outras relações com palavras de sentido próximo quando a dupla procurada não consta do banco de dados. Lamentavelmente não estão disponíveis informações sobre os resultados obtidos com este protótipo.

O UTL (Universal Translation Language) [Fran00] começou em 1999 e ainda está em andamento. Sua meta é definir uma língua adequada para servir como língua intermediária num sistema de tradução automática em dois passos. Esta língua é uma modificação sintaticamente não-ambígua do Esperanto. As modificações propostas no projeto DLT foram tomadas como base, embora haja outras, como uma ordem mais estrita das palavras. Diferentemente da abordagem utilizada no DLT, o usuário do sistema precisaria saber a própria UTL, o que torna o projeto mais voltado para uso profissional. Nenhum protótipo foi disponibilizado ainda.

6 Conclusão

Neste artigo é mostrado como o processamento computacional do Esperanto pode ser otimizado (Seção 5.1) se forem aproveitadas suas regularidades (Seção 4), muito maiores do que as observadas em línguas naturais e em muitas línguas planejadas (Seção 3). Este campo de pesquisa ainda foi muito pouco explorado, e tem um potencial vasto de utilização, como mostrado na Seção 5.2.

Os fenômenos lingüísticos que ocorrem nas línguas naturais também ocorrem em Esperanto [Piro89], tornando-o virtualmente equivalente a elas neste aspecto e justificando seu tratamento computacional. À medida que o movimento

esperantista cresce e se espalha, como vem fazendo desde seu início, tornam-se mais e mais necessários tais tipos de aplicações.

A adoção geral do Esperanto como língua internacional auxiliar é uma solução neutra e democrática para a barreira lingüística presente na internet e no mundo em geral. Não somente soluciona o problema, como o faz de uma forma que exige um esforço pequeno de aprendizado, razoável para todos [Avil01]. O espelhamento de páginas em Esperanto aliado a programas de tradução automática (Seção 5.2.8) também pode auxiliar a resolver este problema, além de ser uma solução de transição até a adoção generalizada do Esperanto ser realizada.

Agradecimentos

A Bei Yi e Ismael Mattos Andrade Ávila pela revisão da primeira versão do texto, que acabou levando a uma grande reestruturação desta versão para a final. A Bei também pelos exemplos em chinês da Seção 3.2.2. A Ismael pela disponibilização da referência [Saus16].

A Termos Lingüísticos

Os termos lingüísticos aparecem abaixo destacados do restante do texto por meio de negrito. Os termos relacionados (de acordo com o nível lingüístico ao qual se aplicam) são explicados juntos.

O conjunto das palavras que constituem uma língua compõem o seu **léxico**. A **fonética** e a **ortografia** tratam este conjunto sob os aspectos sonoro e gráfico, respectivamente. Os sons de uma língua são denominados **fonemas**, e as maneiras gráficas (letras ou ideogramas) de representá-los constituem os **grafemas**. A correspondência fonema-grafema é mais ou menos regular dependendo da língua, podendo ser bastante variada, como visto na Seção 3.2.1. Os fonemas são divididos em **consoantes** e **vogais**, dependendo se durante sua emissão respectivamente ocorre ou não ocorre uma ação de barrar o fluxo de ar. As vogais se distinguem também por poderem formar sílabas por si mesmas. Há línguas que têm associados às vogais **tons** de pronúncia (entonações), que podem servir para distinguir palavras ou para marcar funções gramaticais.

A **morfologia** constitui o estudo de como as palavras se formam, a partir de unidades menores com sentido próprio (**morfemas**). Os morfemas podem ser classificados de acordo com seu papel: as **raízes** são aqueles que têm um papel central no processo de formação da palavra; os **afixos** aparecem em posições periféricas, e só estão presentes em palavras obtidas a partir de outras, mais simples. A morfologia pode tomar como base a fonética ou a ortografia, se o enfoque for a língua falada ou escrita, respectivamente.

A natureza do afixo define dois tipos de processos morfológicos, a **inflexão** e a **derivação**. Na inflexão é adicionado um afixo totalmente geral e que pode ser aplicado a todas (ou quase todas) as palavras de uma classe gramatical; já

a derivação adiciona um afixo que só se junta a um grupo específico de raízes dentro de uma classe. Nestes processos podem ocorrer dois tipos de mudanças nos morfemas para que eles se “encaixem” melhor no conjunto da palavra. Na **alomorfia** o morfema tem alguma parte de sua estrutura sonora ou gráfica alterada. No **truncamento** ocorre o corte de parte da estrutura do morfema, considerada não-essencial.

Os afixos usados no processo de inflexão são comumente chamados de **marcas**. As marcas usadas são divididas em verbais e nominais (associadas a substantivos e também a adjetivos). As marcas mais comuns nas línguas naturais em geral são [Bybe85]:

1. Marcas nominais

- (a) **Gênero:** corresponde à classe lexical —arbitrária ou baseada em qualidades inerentes da entidade referida pela palavra;
- (b) **Número:** refere-se à quantidade (da entidade correspondente à palavra) que está envolvida no contexto;
- (c) **Caso:** mostra o papel semântico que a palavra desempenha na oração completa;

2. Marcas verbais

- (a) **Tempo:** posiciona a ação em relação a um ponto estabelecido no tempo, normalmente o momento da fala;
- (b) **Aspecto:** considera a ação como um todo (aspecto *perfectivo*), normalmente dando ênfase no começo ou no fim, ou como acontecimento em andamento (aspecto *imperfectivo*), de maneira durativa, continuativa ou habitual;
- (c) **Modo:** revela a maneira como o falante apresenta a verdade da proposição no discurso, por exemplo como probabilidade, possibilidade ou certeza;
- (d) **Voz:** indica a perspectiva da qual a ação descrita pelo verbo é vista;
- (e) Concordância de **pessoa** com alguma parte da oração, comumente o sujeito: mostra o papel desta parte na comunicação —a *primeira* pessoa é o falante, a *segunda* o ouvinte, a *terceira* outra pessoa qualquer;
- (f) Concordância de **gênero** ou de **número** com alguma parte da oração, comumente o sujeito.

A **sintaxe** é o estudo de como as palavras formam conjuntos (denominados **sintagmas**) que agem como uma unidade, e de como os sintagmas se relacionam para formar orações. Ela transforma a ordem linear usual das palavras numa ordem hierárquica, na qual a determinação do papel dos sintagmas na oração é realizada.

B Alfabeto Esperantista

Eis as 28 letras que compõem o alfabeto esperantista:

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

Os 28 sons são descritos abaixo agrupados de acordo com sua relação de presença, raridade ou ausência no português [Ribe01]. As letras que os representam são destacadas com negrito.

As vogais são cinco. **A** é pronunciado sempre aberto como em *gato*, nunca fechado como em *cama*. **E** e **o** são sempre fechados como em *cedo*, *povo* — e não abertos como em *seta*, *nota*. **I** e **u** são sempre claros e sonoros como em *grito*, *puro*. É importante ressaltar que **e** e **o** nunca podem ser trocados por **i** e **u**, como acontece na fala comum do português (principalmente no final das palavras), e que as vogais nunca devem ser anasaladas, como em *campo*, *lento*, *ímpar*, *onda*, *muito*.

Existem duas semivogais, **j** e **ŭ**, que correspondem respectivamente ao **i** e ao **u** breves. Estas letras formam ditongos: **aj** (ai em *sai*, *caixa*), **ej** (ei em *lei*, *cheiro*), **oj** (oi em *boi*, *moita*), **uj** (ui em *fui*, *possui*), **aŭ** (au em *autor*, al em *almoço*), **eŭ** (eu em *meu*, *Europa*), **ja** (ia em *modéstia*, *secretária*) etc. Assim a diferença entre os sons de **i** e **j** e de **u** e **ŭ** é que **i** e **u** sempre formam hiatos quando aparecem junto de vogais, enquanto **j** e **ŭ** sempre formam ditongos. Eis exemplos com a pronúncia em português: *feino* = “feino” (fada), *semajno* = “semaino” (semana), *fiulo* = “fiúlo” (homem infame), *maŭro* = “mauro” (mouro).

As consoantes **b**, **d**, **f**, **l**, **m**, **n**, **p**, **t**, **v** e **z** têm a mesma pronúncia em Esperanto e em português. As consoantes restantes dividem-se em dois grupos: o das que têm som equivalente em português e o das que não têm ou cujos sons são pouco comuns. Eis as que têm sons presentes em português: **g** (g em *gato*, *gol*; nunca como o g em *gelo*, *giro*), **ĵ** (j em *jornal*, *janela*), **s** (s em *sapato*, c em *cinema*, nunca com som de z), **ŝ** (ch em *chá*, x em *xarope*) e **k** (c em *casa*, qu em *quente*) e **r** (r em *barata*, *parar*; nunca com som de r em *rico*, *roupa*).

Três consoantes têm sons que raramente ocorrem no português: **c** (ts em *tsar*), **ĉ** (tch em *tchau*) e **ĝ** (dj em *adjetivo*). Outras duas não possuem correspondente: o **h** é um som aspirado (h na palavra inglesa *house*) parecido com o r em *rato*, e o **ĥ** é um som gutural (j na palavra espanhola *mujer*), parecido com o som de limpar a garganta. Este último é um arcaísmo e vem caindo em desuso progressivamente.

A representação dos 28 sons do Esperanto de acordo com o alfabeto fonético internacional é apresentada na tabela abaixo.

a = [a]	b = [b]	c = [ts]	ĉ = [tʃ]	d = [d]	e = [e]	f = [f]
g = [g]	ĝ = [dʒ]	h = [h]	ĥ = [x]	i = [i]	j = [j]	ĵ = [ʒ]
k = [k]	l = [l]	m = [m]	n = [n]	o = [o]	p = [p]	r = [r]
s = [s]	ŝ = [ʃ]	t = [t]	u = [u]	ŭ = [ʊ]	v = [v]	z = [z]

A pronúncia correta das palavras é importante para o entendimento, mas

não é totalmente rígida: é possível variar o local do acento tônico secundário, a pronúncia de algumas letras (por exemplo, e aberto em lugar de e fechado), a duração e o tom das vogais, além de outros fatores considerados secundários, de acordo com o costume da língua de origem do falante.

Quando não é possível representar corretamente as letras **ĉ**, **ĝ**, **ĥ**, **ĵ**, **ŝ** e **ŭ** usa-se a convenção de colocar um **x** (inexistente em Esperanto) à direita da letra que deveria receber o acento: **cx** em vez de **ĉ**, **gx** em vez de **ĝ**, etc. Tal convenção só deve ser usada em caso de real necessidade, por exemplo em mensagens eletrônicas que utilizam o conjunto de caracteres ASCII.

Referências

- [Alle87] James Allen. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, Inc. (1987).
- [Avil01] Ismael Mattos Andrade Ávila. *Acessibilidade em Interfaces Computacionais da Web: Aspectos Lingüísticos*. Trabalho final da disciplina MO622 (Fatores Humanos em Sistemas de Informação), preparado concomitantemente a este artigo.
- [Bick00] Eckhard Bick. *VISL — Flat Structure Parsing*. Disponibilizado em <<http://visl.hum.sdu.dk/visl/eo/parse.html>> (2000).
- [Blan89] Detlev Blanke. *Planned Languages — a Survey of Some of the Main Problems*. Incluído em [Schu89ed].
- [Bybe85] Joan L. Bybee. *Morphology: a Study of the Relation between Meaning and Form*. Typological Studies in Language 9. John Benjamins Publishing Company (1985).
- [Chri01] ChristusRex, Inc. *Pater Noster*. Disponibilizado em <<http://www.christusrex.org/www1/pater/index.html>> (2001).
- [Fran00] Marcos Franco. *UTL — Universal Translation Language*. Disponibilizado em <<http://members.nbci.com/utlpro/utl.htm>> (2000).
- [IPA] *The International Phonetic Alphabet*. Disponibilizado em <<http://mansioningles.metropoli2000.net/images/ipachart.gif>>.
- [Jant93] Pierre Janton. *Esperanto: Language, Literature, and Community*. State University of New York Press (1993).
- [Matt87] Geraldo Mattos. *Gramática Completa do Esperanto*. Fonto (1987).
- [Pass93] José Passini. *Bilingüismo: Utopia ou Antibabel*. EDUFJF, Pontes (1993).

- [Piro89] Claude Piron. *A Few Notes on the Evolution of Esperanto*. Incluído em [Schu89ed].
- [Piro01] Claude Piron. *Esperanto: a Western Language?* Disponibilizado em <http://www.geocities.com/c_piron/westernlanguage.html>, com tradução em português (*Uma Língua Ocidental, o Esperanto?*) em <<http://www.aleph.com.br/kce/artigo25.htm>> (2001).
- [Ribe01] Moisés Braga Ribeiro. *Resumo da Gramática do Esperanto*. Material de apoio não-publicado referente a seminário ministrado na disciplina MO625 (Processamento de Linguagem Natural), ministrada pela professora doutora Ariadne Maria Brito Rizzoni Carvalho no IC/UNICAMP, no segundo semestre de 2000.
- [Sadl89] Victor Sadler. *Knowledge-driven Terminography for Machine Translation*. Incluído em [Schu89ed].
- [Sadl91] Victor Sadler. *Machine Translation Project Reaches Watershed*. *Language Problems and Language Planning* 15 (1991), páginas 78–83.
- [Saus16] René de Saussure. *La Vort-strukturo en Esperanto*. Laborgrupo “Antaŭen”, serio “Brazilaj Kajeroj”, n-ro 4 (original de 1916, republicado em 1985).
- [Schu89a] Klaus Schubert. *Interlinguistics — its aims, its achievements, and its place in language science*. Incluído em [Schu89ed].
- [Schu89b] Klaus Schubert. *An Unplanned Development in Planned Languages*. Incluído em [Schu89ed].
- [Schu89ed] Klaus Schubert (editor). *Interlinguistics: Aspects of the Science of Planned Language*. Mouton de Gruyter (1989).
- [Schu96] Klaus Schubert. *DLT — Faktoj pri Finita Projekto*. Disponibilizado em <<http://esperantic.org/ced/HLT.htm>> (1996).
- [Spro92] Richard Sproat. *Morphology and Computation*. The MIT Press (1992).
- [Wenn01] Bertilo Wennergren. *PMEG — Plena Manlibro de Esperanta Gramatiko*. Disponibilizado em <<http://purl.oclc.org/net/pmeg>> (versão 10.01: 22 de junho de 2001).
- [Wilg00] Gerard van Wilgen. *Ergane*. Disponibilizado em <<http://www.travlang.com/Ergane>>.