

## FERRAMENTA PARA A TRADUÇÃO DA SINTAXE DA LÍNGUA PORTUGUESA PARA A DA LÍNGUA BRASILEIRA DE SINAIS \*

Manoelisa Goebel \*\* Andre Zanki Cordenonsi \*\*\*

### RESUMO

O ser humano e o computador estão cada vez mais próximos, o que é possível através de sistemas computacionais que a Ciência da Computação proporciona. Pode-se, portanto, projetar soluções a fim de solucionar problemas que o homem não havia conseguido resolver utilizando os recursos que dispunha de forma convencional, ou seja, manual. Foi desenvolvida uma pesquisa com alunos da Escola de Ensino Médio Concórdia para Surdos de Santa Rosa-RS, na qual analisaram-se textos redigidos e interpretados por estes, o que possibilitou a análise das diferenças e do grau de dificuldade em termos de estruturação gramatical e sintática devido as diversas diferenças entre a Língua Portuguesa e a LIBRAS. Percebeu-se que, poderia ser utilizado o estudo das Linguagens Regulares para definição e representação de um formalismo. Como o sistema a ser desenvolvido toma por entrada uma cadeia  $x$  e responde “sim”, se  $x$  for uma sentença da linguagem e “não”, em caso contrário, construiu-se um diagrama de transições generalizado chamado de Autômato Finito Determinístico. Desta forma, pretendeu-se desenvolver uma ferramenta que estruturasse gramaticalmente e sintaticamente as frases em português de acordo com a ordem utilizada pelo surdo ao se comunicar através da Língua de sinais brasileira, proporcionando a esta comunidade um recebimento maior das informações, que é necessário para um crescente desenvolvimento social.

**Palavras-Chaves:** Informática na Educação Especial, Língua Brasileira de Sinais, Autômato Finito Determinístico, Classes Gramaticais, Sintaxe, Surdos.

---

\* Trabalho Final de Graduação

\*\* Acadêmica do Curso de Sistemas de Informação

\*\*\* Orientador

## **DEFINIÇÃO DO PROBLEMA**

O ser humano tem, por natureza, a necessidade de se comunicar, e esta comunicação é feita através de um processo de troca e recebimento de mensagens entre duas ou mais pessoas, ou entre dois sistemas. Desta forma, a linguagem é capaz de comunicar uma ampla variedade de idéias e é, talvez o aspecto mais importante que separa os humanos dos outros animais. No entanto, há a necessidade de um meio com dois pólos, um transmissor e outro receptor, onde o meio ou canal pode ser natural (aparelho fonador, sinais gestuais) ou mecânico (imprensa, rádio, televisão, computador pessoal).

Segundo Alzira Costa (COSTA, 1999), a linguagem é definida como um código pelo qual são representadas idéias sobre o mundo através de um sistema convencional de sinais arbitrários utilizados para a comunicação. Desta forma, o código é o meio de representar alguma coisa com outra e linguagem é um meio de representação; a idéia é o conhecimento que o falante necessita ter sobre o objeto; os sistemas são as regras pelas quais os sons são combinados para formar as palavras e as palavras são utilizadas para formar sentenças que representam as idéias.

A língua é um sistema abstrato de regras gramaticais, ou seja, é um conjunto de regras que identificam sua estrutura nos diversos planos (sons, estrutura, formação e classes de palavras, estruturas frasais, semântica, contextualização e uso). Logo, as línguas podem ser orais-auditivas ou espaço-visuais (gestual-visual).

A linguagem utilizada para se comunicar depende do grupo em que está inserido, isto é, depende das especificidades individuais de cada indivíduo. Para os ouvintes, pode-se estabelecer em termos orais-auditivos. No entanto, para os não-ouvintes, pode-se estabelecer em termos gestuais-visuais, onde gestual significa o conjunto de elementos lingüísticos manuais, corporais e faciais necessários para a articulação do sinal.

Existem dois sistemas de comunicação não-verbal: auxiliares e não-auxiliares. Os sistemas não-auxiliares se baseiam em recursos advindos do próprio indivíduo, sem outro suporte além do gesto. Situam-se entre esses a língua oral e a língua de sinais, utilizada por deficientes auditivos. Em oposição, como sistemas auxiliares, estão enquadradas todas aquelas ajudas técnicas que servem de suporte à comunicação. Estas variam de um simples papel e passam às máquinas de escrever, chegando aos computadores e a outras tecnologias relacionadas, tais como os sintetizadores de voz, por exemplo. A oportunidade de se comunicar através

dessas tecnologias torna-se importante para surdos e portadores de deficiência auditiva.

Grande parte do grupo de surdos e deficientes auditivos utiliza a língua de sinais para se comunicar. Ela representa uma modalidade na qual o canal de comunicação é o gestual-visual. Desta forma, seus pensamentos se baseiam na sua língua-mãe que, no Brasil, é denominada de LIBRAS, língua brasileira de sinais. Pelo fato dos usuários de LIBRAS se comunicarem através desta linguagem, percebe-se a diferença em termos sintáticos na estruturação de suas frases, dificultando a compreensão completa dos textos por este grupo.

Para Stokoe, citado por (QUADROS, 1997) entende-se por Língua de Sinais (LS) um sistema lingüístico usado para a comunicação entre pessoas surdas e adquirido como primeira língua por pessoas que não podem ouvir nenhuma língua falada e por filhos de pais surdos.

A LIBRAS, é a língua materna dos surdos brasileiros e ela se torna sua língua-mãe (L1) e a Língua Portuguesa passa a ser sua L2, ou seja, sua segunda língua. E desta forma, é importante ressaltar que este grupo apresenta uma grande dificuldade em aprender uma linguagem escrita que não seja a sua própria. Tal dificuldade não é ocasionada diretamente pela surdez, mas sim porque constroem o pensamento em língua de sinais e precisam escrever em uma segunda língua.

Devido ao fato de não possuírem acesso tão rápido as informações como os ouvintes, este grupo necessita de um meio escrito ou gestual-visual para receber qualquer tipo de informação. Este meio, que pode estar na forma escrita, muitas vezes não passa a informação de uma forma geral, pois está escrito na estrutura da Língua Portuguesa, o que impossibilita sua compreensão total. A sintaxe, que descreve as regras pelas quais as palavras podem ser combinadas em frases gramaticalmente aceitáveis, da Língua Portuguesa difere da sintaxe da língua de sinais escrita e, desta forma, o processo natural de envio e recebimento de informações não é concluído de forma satisfatória.

## OBJETIVOS

O objetivo inicial foi desenvolver uma ferramenta capaz de auxiliar o usuário de LIBRAS na compreensão de textos escritos em Língua Portuguesa, sendo que o problema da compreensão da linguagem falada é um problema de percepção e de difícil solução. No caso da linguagem escrita, este problema é normalmente chamado de compreensão da linguagem natural. O presente trabalho procurou implementar um aplicativo que possibilitasse a passagem de um texto eletrônico, estruturado em

português para a estrutura mais compreensível por aquele que possui como L1 a LIBRAS.

A abordagem tradicional para o desenvolvimento de sistemas de tradução utiliza o PLN (Processamento da Linguagem Natural), o qual busca desenvolver ferramentas capazes de efetuar uma simples correção gramatical entre artigos, assim como total correção sintática de frases envolvendo várias línguas. No entanto, devido à complexidade inerente das gramáticas que são processadas (ex: português-inglês), a fundamentação teórica e prática é extremamente detalhista, focando-se, normalmente, em pequenas parcelas da língua de cada vez.

No entanto, pode-se encontrar na Teoria da Computação a Teoria das Linguagens Formais, que foi desenvolvida na década de 1950 com o objetivo de desenvolver teorias relacionadas com as linguagens naturais, a qual tornou-se importante para o estudo de linguagens artificiais e, em especial, para as linguagens originárias na Ciência da Computação. Atualmente, existem enfoques para aplicações em análise léxica e sintática de linguagens de programação, modelos biológicos, desenho de *hardware* e relacionamentos com linguagens naturais, dando-se ênfase no tratamento de linguagens não-lineares, como planares, espaciais e n-dimensionais.

Em estudo de caso realizado anteriormente (GOEBEL, 2000), percebeu-se que muitas das ferramentas desenvolvidas para surdos têm por finalidade auxiliá-lo com dicionários interativos, que funcionam como um decodificador de sinais baseado no *Sign Writing* (sistema de representação gráfica das línguas de sinais que permite, através de símbolos gráficos, representar as configurações das mãos e seus movimentos, expressões faciais e deslocamentos corporais que são componentes de qualquer língua de sinais (SUTTON, 1997), ou simplesmente no treinamento da voz. Estas ferramentas funcionam como um dicionário eletrônico, sem contribuição para a compreensão da sintaxe do português.

No entanto, foi proposto aqui a modelagem e implementação de uma ferramenta capaz de traduzir textos em português para a estrutura escrita semelhante a LIBRAS. Foi utilizado uma nova abordagem através do estudo das Linguagens Formais e Autômatos e a partir desta teoria, modelou-se um autômato finito determinístico. Um autômato finito pode ser construído como uma função parcial que, dependendo do estado corrente e do símbolo lido, determina o novo estado do autômato.

## LÍNGUA BRASILEIRA DE SINAIS

Segundo (BRITO, 1995), a Língua Brasileira de Sinais, é dotada de uma gramática constituída a partir de elementos constitutivos das palavras ou itens lexicais e de um léxico (o conjunto das palavras da língua) que se

estruturam a partir de mecanismos morfológicos, sintáticos e semânticos que apresentam especificidade, mas seguem também princípios básicos gerais. Estes princípios básicos permitem a criação de um número infinito de construções a partir de um número finito de regras. É dotada também de componentes pragmáticos convencionais, codificados no léxico e nas estruturas da LIBRAS e de princípios pragmáticos que permitem a geração de implícitos sentidos metafóricos, ironias e outros significados não literais. Estes princípios regem também o uso adequado das estruturas lingüísticas da LIBRAS, isto é, permitem aos seus usuários usar estruturas nos diferentes contextos que se lhes apresentam de forma a corresponder às diversas funções lingüísticas que emergem da interação do dia a dia e dos outros tipos de uso da língua.

Em (QUADROS, 1997) cita que nas línguas de sinais, enquanto o emissor constrói uma sentença a partir dos elementos manuais, corporais e faciais, o receptor utiliza os olhos ao invés dos ouvidos para entender o que está sendo comunicado. Desta forma, como a informação lingüística é recebida pelos olhos, os sinais são construídos de acordo com as possibilidades perceptuais do sistema visual humano.

Logo, as relações espaciais nas línguas de sinais são muito complexas. Na LIBRAS, assim como na ASL (Língua de Sinais Americana), as relações gramaticais são especificadas através da manipulação dos sinais no espaço. As relações ocorrem dentro de um espaço definido, na frente do corpo, em uma área limitada pelo topo da cabeça e que se estende até os quadris, sendo que o final de uma sentença na LIBRAS é indicado por uma pausa. A ASL utiliza mecanismos espaciais que fazem com que a informação gramatical se apresente simultaneamente com o sinal. Esses mecanismos envolvem dois aspectos: a incorporação, considerada um mecanismo produtivo na ASL e usada, por exemplo, para expressar localização, número, pessoa e o uso de sinais não-manuais, como movimentos do corpo e expressões faciais. Estes mecanismos também são verificados na LIBRAS.

Exemplos:

PRONOME<sub>c</sub> ENCONTRAR AMIGO

Onde c refere-se a pessoa com quem se está falando.

Interpretações para a sentença:

- Tu encontraste teu amigo? Se o emissor usar a expressão facial interrogativa.
- Tu encontraste O TEU AMIGO. Se salientar o final através de uma expressão facial enfática que pode indicar ironia.
- Tu encontraste teu amigo. Se usar uma expressão facial de naturalidade, simplesmente afirmando.

- Tu não encontraste teu amigo. Se ele usar o movimento da face negando o que está dizendo.

Aqui se verificam as várias possibilidades na LIBRAS, que não estão expressas diretamente nas palavras (sinais), mas na expressão facial usada simultaneamente com o sinal, assim, esses e outros mecanismos espaciais são fundamentais nas línguas de sinais, pois determinam relações sintáticas e semânticas/pragmáticas. No processo de aquisição de língua de sinais, observa-se que os domínios dessas relações são altamente complexos, e que estas relações dependem do domínio do espaço, pois toda língua se processa espacialmente, e especialmente no estabelecimento nominal, o sistema pronominal e a concordância verbal (no caso de verbos que apresentem essa possibilidade).

Assim, os fonemas na LP se estruturam seqüencialmente ou linearmente no tempo enquanto que os “fonemas” da LIBRAS são estruturados simultaneamente ou ao mesmo tempo no espaço. Desta forma, as unidades mínimas se organizam conforme os seguintes parâmetros:

- Configuração de Mãos;
- Ponto de Articulação;
- Movimentação-Orientação;
- Expressão Corporal.

No que diz respeito à ordem das palavras ou constituinte, há diferenças porque o português é uma língua de base sujeito-predicado, enquanto a LIBRAS é uma língua do tipo tópico-comentário. Conforme os exemplos a seguir, pode-se explicar melhor a estruturação das sentenças em LIBRAS. Exemplo:

*O LEÃO MATOU O URSO.*

S P

TODOS OS MENINOS GOSTAM DE FUTEBOL

S P

Pode-se alterar a primeira sentença da seguinte forma:

*O URSO, O LEÃO MATOU OU*

TÓPICO COMENTÁRIO

*AO URSO O LEÃO MATOU*

TÓPICO COMENTÁRIO

Nas frases acima, houve o uso da preposição “a”, mas mesmo assim o urso continuou sendo o objeto direto de “matar” e o leão, seu sujeito. Mas mesmo tendo acontecido a topicalização do objeto, ou seja,

mesmo o objeto sendo o tópico da sentença, o sujeito e o verbo continuaram sendo comentário do corpo.

Esta topicalização é freqüente na fala coloquial, ou seja, em português. No entanto, na LIBRAS, pode-se dizer que é regra geral, ou seja, a ordem TÓPICO-COMENTÁRIO é a preferida, quando não há restrições que impeçam certos constituintes de se deslocarem, mas mesmo assim há um grande número de sentenças SVO. Exemplos:

*VOCÊ LER JORNAL* (= você leu o jornal)

*NÃO ENXERGAR VOCÊ* (= Eu não vi você)

Nestas sentenças, a ordem é SVO, sendo que o sujeito da segunda sentença é um argumento implícito, ou seja, na LIBRAS, assim como na língua portuguesa, o sujeito geral é pressuposto pelo contexto, ou quando se referindo à primeira pessoa é sempre pressuposto como conhecido pelo interlocutor.

Conforme o exposto acima, os usuários de língua de sinais não estão capacitados a produzirem textos em suas línguas naturais, pois as línguas de sinais possuem uma estrutura paralela, com a utilização de gestos complexos que envolvem simultaneamente diversas partes do corpo do sinalizador. Desta forma, a aquisição da Língua Portuguesa pelos surdos é um processo de aquisição de uma segunda língua, a qual é chamada por (QUADROS, 1997) de (L2), o que acaba por dificultar a compreensão de textos escritos na estrutura da Língua Portuguesa.

#### PROCESSAMENTO DE LINGUAGEM

Conforme os fatos descritos, demonstrou-se a problemática relacionada com a compreensão de frases escritas na estrutura da Língua Portuguesa por aqueles surdos que possuem como língua-mãe a língua de sinais brasileira e que devido a ordem gramatical desta não compreendem totalmente um texto. Portanto, pode-se considerar isto, uma problemática relacionada à análise sintática das frases, englobando a Língua Portuguesa e a LIBRAS.

A análise sintática, através do estudo da disposição das frases em uma sentença e de suas relações existentes entre estas, utiliza a decomposição de um período e de cada oração e seus elementos (termos), para verificar a organização e funcionalidade dos termos. Um termo corresponde a um elemento de informação (palavra) e é tratado como unidade funcional da oração, participando da estrutura como um de seus constituintes, chamados de sintagmas (unidades sintáticas), de acordo com suas funções.

Assim, buscou-se estudar a Teoria das Linguagens Formais, que foram desenvolvidas na década de 1950, e que tinham o objetivo de

desenvolver teorias relacionadas com as linguagens naturais. É uma teoria importante para o estudo de linguagens artificiais, como as linguagens originárias na Ciência da Computação. Conforme os estudos foram sendo desenvolvidos, percebeu-se que poderiam aplicá-la em análise léxica e sintática de linguagens de programação, modelo de sistemas biológico, desenhos de circuitos e relacionamentos com linguagens naturais e, atualmente, também tem se utilizado no tratamento de Linguagens Não-Lineares, como Planares, Espaciais e n-Dimensionais. Portanto, através da Teoria das Linguagens Formais, pode-se modelar e desenvolver ferramentas que descrevam sintaticamente a linguagem e seus processos de análise, assim como suas propriedades e limitações algorítmicas.

Em uma linguagem formal, pode-se dizer que um alfabeto é um conjunto finito de símbolos e que um símbolo (ou caractere) é uma entidade abstrata básica que não é definida formalmente (pode-se citar letras e dígitos como exemplos de símbolos freqüentemente usados); e que uma palavra ou cadeia de caracteres ou sentença sobre um alfabeto é uma seqüência finita de símbolos (do alfabeto) justapostos. O tamanho ou comprimento de uma palavra é o número de símbolos que compõem uma palavra, sendo que o prefixo e o sufixo são qualquer seqüência de símbolos iniciais e finais respectivamente das palavras, assim como uma subpalavra de uma palavra é qualquer seqüência de símbolos contígua da palavra.

Logo, uma linguagem formal é um conjunto de palavras sobre o alfabeto, onde há a possibilidade de concatenação, que é uma operação binária sobre a linguagem, que associa a cada par de palavras uma palavra formada pela justaposição da primeira com a segunda.

Conforme (MENEZES, 2000), uma gramática é uma quádrupla ordenada onde:

- $G=(V, T, P, S)$ ;
- $V=$  conjunto finito de símbolos variáveis ou não-terminais;
- $T=$  conjunto finito de símbolos terminais disjuntos de  $V$ ;
- $P=$  conjunto finito de pares, denominados regras de produção, tal que a primeira componente é a palavra de  $(V \cup T)^+$  e a segunda componente é a palavra de  $(V \cup T)^*$ ;
- $S =$  elemento de  $V$  denominado variável inicial;

Para uma regra de produção as regras definem condições de geração das palavras da linguagem. Quando aplicada, uma regra de



produção, esta é denominada de derivação de uma palavra, assim uma aplicação sucessiva de regras de produção permitem derivar as palavras da linguagem representada pela gramática. E desta forma, uma gramática é considerada um formalismo de geração, pois permite derivar (“gerar”) todas as palavras da linguagem que representa.

#### MODELAGEM E FORMALIZAÇÃO

Seguindo um estudo comparativo da estruturação gramatical e sintática conhecida da Língua Portuguesa e da estruturação apresentada pelos surdos ao desenvolverem textos escritos, pretendeu-se encontrar a forma mais comumente utilizada por estes e que não perdesse o sentido do texto ao ser traduzido. Percebeu-se que, para este fim, poderia ser utilizado o estudo das Linguagens Regulares, para que se pudesse realizar uma análise mais aprofundada definindo um formalismo e, conseqüentemente, uma representação.

A modelagem pretendeu possibilitar que uma dada entrada (estado inicial) formada por um texto em estrutura gramatical seguindo a estrutura da Língua Portuguesa tivesse como saída (estado final) o texto em estrutura escrita pelo surdo seguindo a Língua Brasileira de Sinais, sendo que esta tradução é geralmente chamada de *glosa*.

Como se trata de um texto, cita-se o caso de analisadores léxicos e processadores de textos (como também algumas ferramentas de processadores de textos), onde se têm exemplos que se enquadram dentro de um *Sistema de Estados Finitos*, como o desenvolvido por (LUCCHESI, 1992), que são associados a diversos tipos de sistemas naturais, pois cada estado é capaz de memorizar a estrutura do prefixo da palavra em análise. Assim, existem entradas e saídas discretas, podendo haver um número finito e pré-definido de estados, sendo que cada estado resume somente as informações do passado necessário para determinar as ações para a próxima entrada.

O funcionamento de um autômato finito pode ser explicado da seguinte forma, o qual foi a base para o desenvolvimento da modelagem e implementação da ferramenta. O sistema a ser desenvolvido toma por entrada uma cadeia  $x$  e responde “sim” se  $x$  for uma sentença da linguagem e “não” em caso contrário. Desta forma, para compilar expressões regulares em um reconhecedor é preciso a construção de um diagrama de transições generalizado chamado de autômato finito.

Um Autômato Finito Determinístico (AFD) é visto como uma máquina composta por três partes:

- Fita: dispositivo de entrada que contém a informação a ser processada;

- Unidade de Controle: possui uma unidade de leitura (cabeça da fita), que processa uma célula da fita de cada vez e movimenta-se exclusivamente para a direita, sendo encarregada de refletir o estado corrente da máquina;
- Programa ou Função de Transição: comanda as leituras e define o estado da máquina.

Um AFD possui, no máximo, uma transição a partir de cada estado para qualquer símbolo de entrada. Logo, para cada entrada na tabela de transições existe um único estado final. Conseqüentemente, é mais fácil determinar se um autômato finito determinístico aceita uma cadeia de entrada, dado que existe no máximo um único percurso, rotulado por aquela cadeia, a partir do estado inicial.

A Língua Brasileira de Sinais não possui uma estruturação gramatical definida, sendo que são poucos os lingüistas que procuraram estudá-la até hoje, podendo-se citar entre estes, Lucinda Ferreira Brito (Por uma Gramática de Sinais), doutora em lingüística da UFRJ, a qual cita que muitos dos surdos utilizam a topicalização, ou seja, a estrutura de suas frases é do genérico para o específico, mas não há como definir qual a ordem de classes gramaticais será utilizada.

Para poder definir um autômato, necessita-se definir seu estado, evento, ações, condições, transições. No caso das classes gramaticais, teremos o seguinte:

- Estado: situação de um objeto em determinado momento.
- Estado Inicial: classe gramatical conforme a Língua Portuguesa;
- Estado Final: após a transição, existência ou não desta classe, que poderá ser total ou parcial;
- Eventos: algum motivo que desencadeia a transição. Neste caso, quando é solicitada a tradução do texto, ou seja, quando o texto eletrônico em Língua Portuguesa será passado para a estruturação gramaticalmente compreensível pelos surdos através de uma solicitação;
- Condição: expressão booleana avaliada quando um evento ocorre. Se for satisfeita, a transição ocorre. Quando dependem de uma condição, as transições são chamadas de transições guardadas. Assim, pode-se dizer que, após a solicitação da tradução do texto, serão verificadas as condições de existência para cada uma das classes gramaticais que compõem o texto. Ou seja, neste novo estado, após a transição, muitas delas poderão aparecer

parcialmente assim como outras não terão sentido em permanecer, pois não são utilizadas ou até desconhecidas pelos surdos;

- Transições: um objeto passa de um estado para outro quando um evento ocorre e a ligação ou relacionamento entre estes dois estados é chamada de transição. Ela poderá ser da seguinte forma:
  - Efetivada sem restrição;
  - Efetiva através de uma condição (chamada de Transição Guardada);
  - Não ter nenhuma ação associada a ela;
  - Ter ações associadas a ela e que não irão interferir em seu destino; seu destino poderá ser determinado pelo resultado de uma ação associada a ela.

Para que o texto eletrônico em português seja traduzido para a estrutura desejada, a qual é chamada de *glosa*, há a necessidade de uma transição ou seja, de uma passagem que precisará seguir alguns parâmetros para que o estado final seja o definido.

Neste caso, o novo estado do texto eletrônico desejado será aquele capaz de passar a mesma idéia que havia no texto em português, mas seguindo a estrutura definida para o estado final como sendo a mais utilizada pelos surdos. Em se tratando de classes gramaticais, algumas permanecerão da mesma forma, outras serão modificadas e algumas não existirão.

Conforme (MENEZES, 2000), a representação de um Autômato Finito Determinístico (AFD) também poderá ser da seguinte forma, como uma 5-upla:

$$M = (\Sigma, Q, \delta, q_0, F)$$

Onde:

- $\Sigma$  = alfabeto de símbolos de entrada;
- $Q$  = conjunto de estados possíveis do autômato, o qual é finito;
- $\delta$  = função programa ou função transição, que é uma função parcial;
- $q_0$  = estado inicial tal que  $q_0$  é elemento de  $Q$ ;
- $F$  = conjunto de todos os estados possíveis tal que  $F$  está contido em  $Q$ .

Desta forma, a formalização da modelagem proposta, pode seguir a seguinte estrutura:

- $C_p$ : conjunto das classes gramaticais da Língua Portuguesa;
- $C_s$ : conjunto das classes gramaticais compreensíveis e utilizadas pelos surdos;
- $f$ : função parcial;
- T: texto eletrônico inicial em português.

A função transição( $f$ ) para (MENEZES,2000) é uma função parcial, pois é uma relação onde cada elemento do domínio está relacionado com, no máximo, um elemento do contra-domínio, ou seja, para cada estado inicial de cada classe gramatical, existirá somente um estado final, que será aquele que é usado e compreendido pelos surdos. Assim, para cada palavra de entrada, o processamento do autômato finito efetua uma aplicação sucessiva da função programa sobre cada símbolo desta, até uma condição de parada aparecer. Existirá portanto, uma linguagem aceita pelo Autômato Finito, a qual será o conjunto de todas as palavras pertencentes ao alfabeto inclusive as vazias, que foram aceitas pela gramática, no caso, o formado pelas classes gramaticais usadas e com suas variações conforme a utilização dos surdos.

Conforme a pesquisa realizada, existiram casos de palavras que não foram aceitas pela gramática, ou seja o autômato não aceitará as classes gramaticais e suas variâncias não utilizadas, formando um conjunto de palavras rejeitadas.

No caso do alfabeto de símbolos de entrada( $\Sigma$ ), aqui será o alfabeto da Língua Portuguesa; o conjunto de estados possíveis(Q) será formado pelas classes gramaticais da Língua Portuguesa( $C_p$ ); o conjunto de todos os estados possíveis(F) é aquele representado pelas as classes gramaticais usadas e compreendidas pelos surdos com suas respectivas variações( $C_s$ ), caso ocorram.

Como algumas classes gramaticais foram modificadas e/ou eliminadas, a função de transição pode ser definida da seguinte forma:

$$F: f(C_p) \rightarrow \gamma, \text{ onde } \gamma \in C_s.$$

Desta forma, uma ou mais funções geraram resultados  $\gamma$ , os quais deveriam pertencer ao conjunto  $C_s$ , isto é, os resultados gerados a partir desta função pertencem ao conjunto formado pelas classes gramaticais usadas e compreendidas pelos usuários de LIBRAS. Por isso, para cada classe  $C_p$  houve uma ou mais subfunções.

## COMPARAÇÃO ENTRE LP E LIBRAS

Buscou-se através de estudo comparativo entre os testes aplicados e gramática de Língua Portuguesa conforme (CEGALLA, 1996) proporcionar uma comparação entre esta gramática e a da LIBRAS.

Para esta avaliação, conforme estudo realizado anteriormente (GOEBEL, 2001), tem-se as seguintes estruturas para as classes gramaticais que aparecerão no texto final. Assim, as classes receberam o nome de *nome\_classe gramatical\_glosa*, pois terão contidas as palavras que retornarão nas frases. A seguir segue o estudo entre estas duas línguas e o resultado final deste.

### Adjetivo

Palavras que servem para expressar as qualidades ou características dos seres, possuem importante papel no sentido de suas frases e ainda podem exercer funções sintáticas de predicativo e adjunto adnominal.

Conforme (FELIPE, 1997), geralmente na LIBRAS, vem após o substantivo que o qualifica.

Adjetivo\_glosa: considerados os mais usados e compreensíveis pelo grupo, como por exemplo, “os tons de cores”, “bonita”, “bonito”, “lindo”, “linda”, “triste”, “alegre”, “mais do que”, “menos do que”, “igual” entre outros.

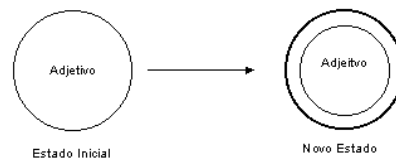


Figura 1- Autômato Adjetivo\_glosa

### Advérbio

Podem modificar o sentido do verbo, do adjetivo e do próprio advérbio.

Advérbio-glosa: como é necessário para as frases não perderem seu sentido real, foram somente considerados aqueles que passam a noção de tempo, lugar e intensidade, como por exemplo, ontem, hoje, cedo, demais, etc.

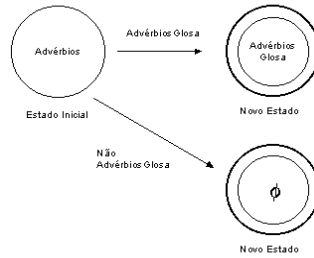


Figura 2- Autômato Advérbio\_glosa

## Artigo

Classe anteposta ao substantivo e que permite dar sentido determinado ou indeterminado aos seres, indicando também tempo, gênero e número dos substantivos.

Artigo\_glosa: percebeu-se o uso dos artigos definidos, pois são utilizados principalmente para indicar a pessoa de quem estão falando o referenciando o assunto em questão no diálogo.

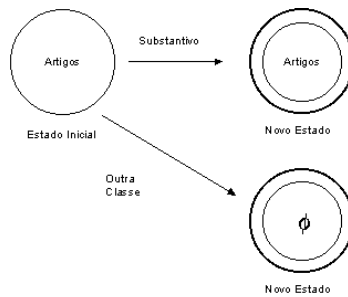


Figura 3- Autômato Artigo\_glosa

## Conjunção

Utilizada para ligar orações.

Conjunção-glosa: muitas vezes foram utilizada incorretamente na estrutura da frase retorno outro sentido ou até desconfigurando o sentido desta, mas no entanto para dar consistência ao texto, serão retornadas as mais usadas como, “*mas*”, “*e*”, “*também*”, “*ou*”, “*antes*”, “*que*”, “*porque*”, “*como*”, “*se*”.

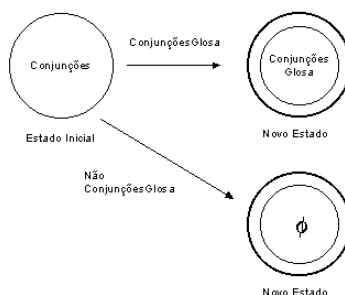


Figura 4-Autômato Conjunção\_glosa

## Preposição

Considerada na gramática como uma palavra invariável que liga um termo dependente a um termo principal, estabelecendo uma relação entre ambos.

No caso do uso das preposições pelo grupo avaliado, constatou-se o uso incorreto desta classe na maioria todos os textos, muitas vezes sendo colocados desnecessariamente. Percebeu-se que para eles não há sentido esta classe, mas se no resultado final da tradução não houver esta classe poderá a frase perder seu sentido, sendo necessário considerar as mais usadas por nós ouvintes e compreendidas pelo surdo.

Preposição-glosa: como explicado anteriormente, considerou-se apenas as seguintes: “*antes*”, “*após*”, “*até*”, “*com*”, “*como*”, “*para*”, “*depois*”, pois dão sentido de ligação temporal.

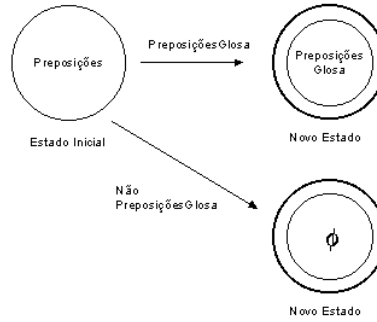


Figura 5- Autômato Preposição-glosa

### Pronome

Permite representar ou determinar o nome dos seres indicando a pessoa do discurso. Segundo (FELIPE, 1997), quando uma pessoa surda está conversando, ele pode omitir a primeira pessoa e a segunda porque, pelo contexto, as pessoas que estão interagindo sabem a qual das duas o verbo está relacionado. Por isso, quando essas pessoas estão sendo utilizadas pode ser para dar ênfase á frase. Geralmente para indicar uma pessoa no singular, é utilizado a configuração predominante em “d” (dedo indicador estendido), que difere da orientação do sinal “eu”, que aponta para o peito e do sinal você que aponta para o receptor e o sinal para ele/ela é apontar para uma pessoa que não está na conversa ou para um lugar convencionalizado para uma terceira pessoa que está sendo mencionada.

Pronome-glosa: desta forma foram considerados todos os pronomes pessoais exceto “tu” e “vós”, pronomes demonstrativos como “este”, “esta”, “aquele”, “aquela”, “quem”, entre outros, que dão sentido a frase e que também mesmo não sendo utilizado pelo surdo é compreensível por este.



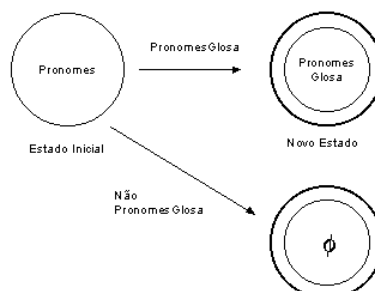


Figura 6-Autômato Pronomes\_glosa

### Substantivos

Foram consideradas todas as palavras integrantes do texto que não se enquadraram em nenhuma das outras classes, pois é a classe responsável por designar os seres, sendo uma classe muito extensa. A condição para esta classe definida é a ocorrência de artigos antes da palavra.

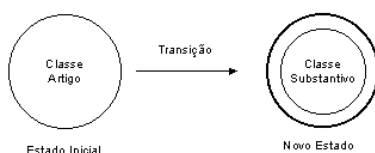


Figura 7-Autômato Ocorrência Substantivo

### Verbos

São capazes de permitir a expressão de ação, estado, fato ou fenômeno. Na LIBRAS, basicamente existem dois tipos de verbo:

- a) verbos que não possuem marca de concordância, embora podendo ter flexão para o aspecto verbal;
- b) verbos que possuem marca de concordância.

O surdo, quando usa língua de sinais para se comunicar, tem o sinal para a ação (verbo), mas não há como fazer um sinal para cada tempo verbal, com seu modo. Desta forma, foi definido um autômato, onde o verbo conjugado do texto em português terá como estado final sua conjugação no infinitivo.

A Língua Portuguesa possui mais de 11.000 verbos, no entanto foram classificados cerca de 183 em ordem de terminação verbal. Existem

os verbos regulares, os irregulares e os defectivos, como estes últimos ou modificam seu radical verbal ou possuem somente algumas conjugações, não houve tempo para classificação de todos, sendo estes não modificados na sua estrutura final.

Assim os verbos foram classificados conforme seus radicais verbais posteriormente conforme sua vogal temática, passados para o infinitivo de acordo com suas terminações (ar, er, ir).

### **IMPLEMENTAÇÃO DA FERRAMENTA**

Implementar uma ferramenta baseada em autômatos finitos significa, que deverá haver uma entrada com um determinado estado e uma saída com um outro determinado estado.

Para isso buscou-se através dos testes aplicados e da bibliografia disponível a respeito da gramática e sintática da LIBRAS definir quais seriam os dados que poderiam permanecer no resultado final e quais seus estados.

Desta forma, pode-se dizer que em relação às classes gramaticais, as palavras foram armazenadas de acordo com sua classificação e cada classe recebeu um código para posterior comparação entre elas.

Para que possa ser possível verificar quais palavras retornarão no final da tradução e quais não, foi criada uma função (chamada *nome\_classe gramatical\_glosa*), onde somente retornará verdadeira aquela palavra que estiver contida nela de acordo com sua classe a qual faz comparação, sendo que as outras possuirão um retorno falso, ou seja, vazio, não permanecendo no estado final do texto.

Em relação a classe gramatical dos verbos, houve a necessidade de um estudo mais detalhado, pois se dividi em radical verbal, vogal temática e terminações verbais de acordo com sua conjugação verbal a qual faz parte (AR- ER- IR).

Como os radicais verbais forma armazenados de acordo com sua conjugação, através de uma verificação da existência do prefixo da palavra em questão no arquivo e de sua terminação verbal presente na função é feita a passagem para a forma nominal infinitivo e conforme o resultado dos testes aplicados anteriormente, grande parte do grupo analisado, utiliza esta forma ao escrever o verbo na frase, pois não existe uma forma de expressar para cada pessoa, em língua de sinais, os verbos conjugados.

A classe gramatical dos verbos, foi a que requereu mais tempo, pois existem mais de 11.000 verbos na nossa Língua Portuguesa e foi preciso pesquisar os verbos mais usados no nosso dia, a fim de proporcionar que o texto final continuasse passando a mesma informação que o texto inicial, ou seja, não perdesse consistência dos seus dados, sendo que

algumas vezes poderá não ocorrer a troca pois existem muitos radicais não armazenados.

Cada palavra da frase analisada é verificada com os elementos integrantes das classes gramaticais armazenadas, e conforme o código pertencente a esta é comparada, por radical, caso seja verbo ou por comparação total desta, caso seja qualquer outra classe.

Outras questões referentes aos verbos, são as suas divisões em regular, irregular e defectivo, o que quer dizer que os verbos irregulares por sofrerem alteração no seu radical e/ou terminação verbal e os anômalos por não possuírem conjugação completa e não usarem todos os modos verbais, não serão considerados nesta versão do aplicativo, o que fará com que em alguns textos a tradução permaneça quase que sem alterações.

Em relação as classes das preposições e conjunções houve a necessidade de se criar a função glosa, pois mesmo que os surdos não utilizem-nas para conversar ou escrever, há a necessidade de constarem nos textos finais devido ao fato destes ficarem algumas vezes sem sentido.

A ferramenta faz a comparação sílaba por sílaba de cada frase do texto, sendo que aquelas palavras que forem consideradas glosa permanecerão no estado final e os verbos passarão para a forma nominal no infinitivo.

No caso da classe substantivo, considerou-se que aquelas palavras comparadas e não encontradas tanto nas classes arquivadas como nas respectivas funções glosa, seriam substantivos podendo retornar no estado final do texto, pois esta classe é aquela responsável por designar os seres, o que quer dizer que dependendo da classe que a anteceder ou do contexto da frase, ela poderá ser ou não um substantivo, o que seria mais complexo para implementar.

Em relação à ordenação das frases, como são poucos os estudiosos que procuram pesquisar a respeito deste assunto, as alterações implementadas estão em relação a colocação dos verbos e conjunções, substantivos e adjetivos e verbos e adjetivos, as quais foram analisadas com os textos redigidos pelo grupo avaliado.

## CONCLUSÕES

Desenvolver um projeto de pesquisa voltado à comunidade surda requer muita atenção a determinados fatos como: não são todos os surdos que utilizam a língua de sinais brasileira e portanto, sua estrutura de comunicação e escrita será diferente. É importante ressaltar aqui, que esta ferramenta foi desenvolvida para aqueles que têm como L1 a LIBRAS, mas isto também não quer dizer que se pode generalizar que todos tenham dificuldade de compreender os textos na estrutura da LP, assim como não pode-se generalizar que todos os brasileiros não compreendem determinados textos devido a vocabulário fraco.

Outro fato importante de ser relatado é em relação às interpretações de textos feitas pelo grupo avaliado, muitos não conseguiam compreender o texto não por causa somente da estrutura em português mas também devido ao fato de não conhecerem alguns dos vocábulos que ali constavam .

Para chegar a conclusão da necessidade de uma ferramenta que fizesse a tradução, foi realizado anteriormente um estudo de caso (GOEBEL,2000), no qual verificou-se que a maioria das ferramentas desenvolvidas para os surdos tinham por finalidade auxiliá-lo com dicionários interativos, que funcionavam como um decodificador de sinais baseado no *Sign Writing* (sistema de representação gráfica das línguas de sinais), que permite, através de símbolos gráficos, representar as configurações das mãos e seus movimentos, expressões faciais e deslocamentos corporais que são componentes de qualquer língua de sinais ou simplesmente sistemas para treinamento da voz. Estas ferramentas funcionam como um dicionário eletrônico, sem contribuição para a compreensão da sintaxe do português.

Este trabalho procurou encontrar, através do estudo das Linguagens Formais, uma forma de solucionar o problema da não total compreensão de textos pelos surdos, sendo que tal fato ocorre devido a língua de sinais ser uma língua baseada no uso de espaço e movimentos das mãos, no lugar de modulações do trato vocal. Desta forma, o sistema gramatical é mais complexo, fazendo uso de padrões e contrastes espaciais.

Em relação à sintaxe, conforme (PEREIRA,1990) na língua de sinais americana, o uso do espaço parece ter papel importante, pois dentro do “espaço sinalizado” (a área em que os sinais são feitos), os sinais podem ser movidos de uma localização para outra, indicando diferenças no sujeito e no objeto. Isto significa que se pode estabelecer localizações no espaço para dois personagens. Por exemplo, mover a mão relacionada a um destes personagens para o outro, indicando algo, ou seja, uma relação de receptor e

transmissor da mensagem. No entanto, este fato pode ficar subtendido em textos escritos em português.

Procurou-se desenvolver uma ferramenta capaz de possibilitar àquele surdo ou deficiente auditivo que possui como língua-mãe a Língua Brasileira de Sinais uma maior aquisição das informações eletrônicas, pois o computador é um dos meios que mais possibilita a recepção visual destas informações. E como possuem a Língua Portuguesa como L2, não compreendem totalmente um texto estruturado nesta língua, geralmente acarretando dúvidas em relação ao seu conteúdo.

Devido a pouca compreensão dos textos ocasionada pela estrutura sintática das frases em português, modelaram-se autômatos, os quais definiram os estados finais de cada classe gramatical encontrada nos textos eletrônicos.

A grande dificuldade encontrada foi a de determinar a ordem gramatical utilizada pela maioria dos surdos. Em relação à sintaxe, como citado em Pereira, esta tem tido atenção de estudiosos. No entanto, isto ocorre em relação à língua americana de sinais. Em relação à línguas de sinais usadas no Brasil, vários trabalhos estão sendo realizados nesta área da lingüística.

Há portanto a necessidade de continuação das pesquisas nas áreas envolvidas neste trabalho e há também a possibilidade de implementação futura de ferramentas mais completas que permitam a total tradução dos textos eletrônicos de acordo com a estrutura da língua de sinais utilizada pela grande maioria dos surdos e deficientes auditivos.

Como o surdo possui mais desenvolvido o sentido da visão e sabe-se que o computador é uma máquina capaz de permitir que as informações cheguem até as pessoas instantaneamente devido o avanço tecnológico, que popularizou a *internet* e, conseqüentemente, difundiu o uso dos textos em formato eletrônico, esta máquina torna-se ferramenta essencial na popularização do conhecimento. Desta forma, pretende-se em estudos futuros definir uma estrutura sintática das frases em português, que proporcione à comunidade surda um recebimento maior das informações, que é necessário para um crescente desenvolvimento social.

É importante salientar que a língua de sinais brasileira é muito rica, com uma grande variedade de recursos. Portanto, abre-se a possibilidade de desenvolver futuras investigações, a fim de gerar e aperfeiçoar novas ferramentas, proporcionando também, a esta comunidade, o acompanhamento deste avanço.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BRITO, Lucinda Ferreira. 1995. **Por uma Gramática de Sinais**. Rio de Janeiro: Tempo Brasileiro.
- CEGALLA, Domingos Paschoal . 1996. **Novíssima Gramática da Língua Portuguesa**. 39ª Edição. São Paulo: Editora Nacional.
- COSTA Nobre, Maria Alzira; RAMPELOTTO, Elisane. 1999. **Curso de Capacitação de Professores para Atuação com PNEEs, visando a Educação Inclusiva**. Santa Maria – RS.
- FELIPE, Tanya A. 1997. **Introdução à Gramática da LIBRAS. Série Atualidades Pedagógicas**. Volume III. Brasília. SEESP.
- GOEBEL, Manoelisa. 2000. **Um Estudo para Auxiliar Usuários de LIBRAS**. Anais IV SEPE- Centro Universitário Franciscano. Santa Maria – RS.
- GOEBEL, Manoelisa. 2001. **Proposta de um Autômato Finito Determinístico para a Glosa Português – LIBRAS**. Anais do XII Simpósio Brasileiro de Informática na Educação - SBIE . Vitória – ES.
- LUCCHESI, Cláudio L.; KOWALTOWSKI, Tomasz. **Applications of Finite Automata Representing Large Vocabularies**. 1992, Relatório Técnico DCC-01/92. Departamento de Ciência da Computação, Universidade Estadual de Campinas – SP.
- MENEZES, Paulo Fernando Blauth. 2000. **Linguagens Formais e Autômatos**. 3ª Edição. Porto Alegre: Editora Sagra - Luzzatto.
- PEREIRA, Maria Cristina da Cunha. **Aspectos Sintáticos da Língua Brasileira de Sinais**. 1990.
- QUADROS, Ronice Müller . 1997. **Educação de Surdos- A aquisição da Linguagem**. 1ª Edição. Porto Alegre: Artes Médicas.
- SUTTON, Valerie. 1997. **Sign Writing Site**. Disponível por http: // [www.signwriting.org](http://www.signwriting.org).