CORDENONSI, A. Z. ; GOEBEL, M. . Proposta de um Autômato Finito Determinístico para a Glosa Português-Libras. In: XII Simpósio Brasileiro de Informática na Educação, 2001, Vitória - ES. Anais do XII Simpósio Brasileiro de Informática na Educação. Vitória : Crediné Silva de Menezes, Davidson Cury, Orivaldo de Lira Tavares, 2001. v. 1. p. 275-284.

PROPOSTA DE UM AUTÔMATO FINITO DETERMINÍSTICO PARA A GLOSA PORTUGUÊS – LIBRAS

RESUMO: Grande parte dos surdos utiliza a Língua Brasileira de Sinais (LIBRAS) para a sua comunicação. No entanto, alguns encontram dificuldade para compreender totalmente um texto, pois o português não é a sua língua–mãe, e possui uma ordem diferente da definida pela LIBRAS. Para minimizar este problema, foi desenvolvido um Autômato Finito Determinístico para a transformação de textos eletrônicos em português para textos no formato compreensível por este grupo, o que é denominado Glosa Português–LIBRAS.

PALAVRAS – CHAVES: Informática na Educação Especial, Aspectos Sociais de Informática na Educação, Língua Brasileira de Sinais, glosa, Autômato Finito Determinístico.

Manoelisa Goebel
(manoelisa@hotmail.com)

Andre Zanki Cordenonsi
(andrezc@unifra.br)

Centro Universitário Franciscano (UNIFRA)
Rua dos Andradas, 1614 -Santa Maria – RS
CEP 97010-032
+55 (55) 220-1200

PROPOSTA DE UM AUTÔMATO FINITO DETERMINÍSTICO PARA A GLOSA PORTUGUÊS – LIBRAS

1 Introdução

A linguagem utilizada para comunicação depende do grupo em que se está inserido, ou seja, depende das especificidades individuais de cada indivíduo. Para os ouvintes, pode-se estabelecer em termos orais-auditivo. No entanto, para os não ouvintes (surdos), 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.

Grande parte do grupo dos surdos e deficientes auditivos utilizam a língua de sinais, a qual representa uma modalidade onde 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 da língua de sinais se comunicarem através desta, a estruturação das frases na Língua Portuguesa, ou melhor, a sintaxe da Língua Portuguesa (a qual se torna sua segunda língua) difere da sintaxe da língua de sinais brasileira escrita, dificultando sua compreensão completa de textos.

Devido ao fato de não possuírem um acesso tão rápido as informações como os ouvintes, os surdos necessitam de um meio escrito ou gestual-visual para receberem qualquer tipo de informação. Quando este meio está na forma escrita, muitas vezes a informação não é compreendida de uma forma geral, pois está na forma escrita da Língua Portuguesa, o que impossibilita sua compreensão total.

Em Quadros [1], é citado 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, 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. O mesmo processo acontece com a *American Signal Language (*Língua de Sinais Americana), a qual utiliza mecanismos espaciais que fazem com que a informação gramatical se apresente simultaneamente com o sinal. Estes mecanismos envolvem quatro aspectos: localização, número, pessoa e o uso de sinais não-manuais, como os movimentos do corpo e expressões faciais. Estes mecanismos também são verificados na LIBRAS.

As línguas de sinais, assim como as orais, possuem uma gramática própria com regras específicas em todos os níveis lingüísticos: morfológico, fonológico, sintático e semântico. Os fonemas no português se estruturam seqüencialmente ou linearmente no tempo, enquanto que os "fonemas" da LIBRAS são estruturados simultaneamente no espaço. Desta forma, as unidades mínimas se organizam conforme os seguintes parâmetros: (a) Configuração de mãos; (b) Ponto de articulação; (c) Movimentação – Orientação; e (d) Expressão corporal.

No que diz respeito à ordem das palavras ou constituintes, Brito [2] afirma que há diferenças, pois o português é uma língua de base Sujeito–Predicado, enquanto a LIBRAS é uma língua do tipo Tópico–Comentário. Esta topicalização é freqüente na fala coloquial dos ouvintes. No entanto, na LIBRAS, pode-se dizer que é regra geral, onde a ordem Tópico–Comentário é a preferida quando não há restrições que impeçam certos constituintes de se deslocarem. Mesmo assim, há um grande número de sentenças utilizando a ordem Sujeito–Verbo– Objeto (SVO).

Conforme o exposto acima, os usuários de língua de sinais não estão capacitados a produzirem textos em suas línguas naturais, pois estas 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 [1] de L2, o que acaba por dificultar a compreensão de textos escritos na estrutura da Língua Portuguesa.

Será demonstrada aqui a modelagem de uma ferramenta capaz de traduzir textos em português para a LIBRAS, de uma forma geral e completa. Como a estrutura desta língua é mais simples do que as gramáticas das linguagens orais, será utilizada uma abordagem através do estudo das Linguagens Formais. A partir desta teoria, modelou-se um autômato finito determinístico que possa realizar a tradução seguindo a forma supostamente mais correta que respeite a cultura surda.

A modelagem proposta pretende possibilitar que uma dada entrada (estado inicial) formada por um texto em estrutura gramatical seguindo a estrutura da Língua Portuguesa, tenha como saída (estado final) um texto com uma estrutura semelhante à utiilzada pela Língua Brasileira de Sinais, sendo que esta tradução, ou passagem de um estado para outro, é comumente denominada de *glosa*.

2 Estudo da Gramática e Sintaxe da LIBRAS

Foi desenvolvido um estudo da cultura e identidades surdas através de uma pesquisa realizada com alunos da Escola de Ensino Médio Concórdia para Surdos de Santa Rosa-RS, onde analisaram-se textos redigidos por estes.

Verificou-se a estruturação das frases em seus textos, o que possibilitou a análise do grau de dificuldade em termos de estruturação gramatical e as diversas diferenças entre a língua portuguesa e a LIBRAS, em relação à lingüística e estrutura sintática das mesmas. Foram localizados pontos críticos que impedem a total compreensão de uma informação escrita na estrutura da Língua Portuguesa pelos surdos como, por exemplo, uso dos tempos verbais.

Na figura 1, são apresentados quatro gráficos que exemplificam o estudo realizado, demonstrando os resultados para as classes adjetivo e verbos. Foram analisados textos de trinta e dois alunos no total, sendo quinze do primeiro grau e dezessete do segundo grau.

A principal dificuldade nesta fase foi determinar a ordem gramatical utilizada pela maioria do grupo pesquisado, pois a LIBRAS utiliza outros padrões como, por exemplo, o uso do espaço e movimentos das mãos no lugar de modulações do trato vocal. Este fato é referenciado em Brito [2], onde o autor afirma que os surdos preferem, muitas vezes, utilizar a topicalização para estruturar suas frases, ao contrário do grupo de ouvintes, onde a maioria das frases é construída seguindo a ordem Sujeito–Verbo–Objeto. Ou seja, a ordem das frases da grande maioria dos surdos parte do genérico para o específico, seguindo a ordem que utilizam ao se comunicarem por sinais.

Desta forma, seguindo os testes aplicados, pode-se modelar a ferramenta proposta através de dois processos. O primeiro será definido sobre as classes gramaticais encontradas no texto. Algumas poderão sofrer mudanças e outras, muitas vezes, não serão encontradas na fase final. O segundo processo será responsável pela passagem do texto eletrônico para a estrutura utilizada pela grande parte da comunidade surda, ou seja, seguirá uma nova ordem sintática estabelecida através dos testes. Esta nova ordem sintática ainda encontra-se em fase de estruturação.

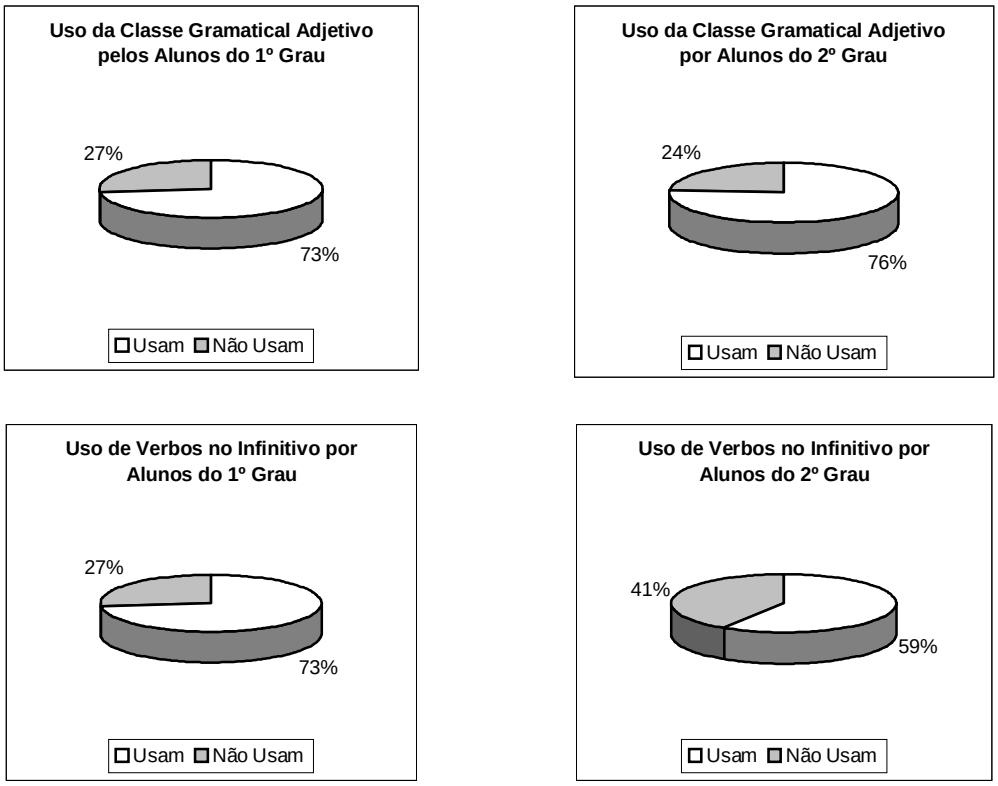


Figura 1- Percentual de utilização das classes gramaticais Adjetivo e Verbo

3 Definição de um Autômato Finito Determinístico

Seguindo um estudo comparativo da estruturação gramatical e sintática conhecida da Língua Portuguesa, baseado-se em Cegalla [3], e da estruturação apresentada pelos surdos ao desenvolverem os textos escritos, procurou-se encontrar a forma mais comumente utilizada por estes. Percebeu-se que, para este fim, poderia ser utilizado o estudo das Linguagens Regulares para uma definição e representação em um formalismo.

Conforme Vicencio [4], uma linguagem pode ser considerada como um conjunto cujos elementos são precisamente definidos através de regras e, para descrevê-la, há a necessidade de um sistema formal que defina os elementos deste conjunto. Deve-se definir os elementos do conjunto e, para cada um deles, associar uma estrutura, sendo que tais sistemas formais são denominados gramáticas. Assim, pode-se ter associada à gramática uma ou mais estruturas, onde cada uma possui um significado próprio, como no caso de Lucchesi [5]. Caso existam várias estruturas relacionadas a uma mesma sentença, ocorre o que se denomina de "ambigüidade", e é função da gramática fazer previsões corretas sobre este tipo de ambigüidade estrutural.

Pesquisas desenvolvidas utilizando-se autômatos já foram desenvolvidas. Dentre estas, pode-se destacar a de Lucchesi [5], a qual procurava modelar ferramentas que fossem capazes de fornecer, através da construção parcial de um autômato finito determinístico, uma representação de um vocabulário a partir de uma linguagem natural.

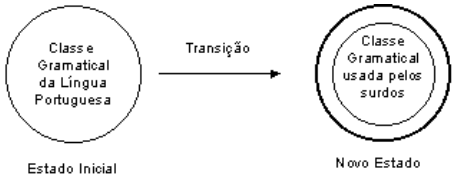
Para o desenvolvimento desta proposta, houve a necessidade de se encontrar um reconhecedor para a linguagem. Como o sistema a ser desenvolvido toma por entrada uma cadeia x e responde "sim", se x for uma sentença da linguagem e "não", em caso contrário, construiu-se um diagrama de transições generalizado chamado de Autômato Finito Determinístico (AFD), conforme Menezes [6].

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.

Para que o texto eletrônico em português seja traduzido para a estrutura desejada, a qual é chamada de *glosa*, necessitou-se de uma transição, ou seja, de uma passagem que seguisse certos parâmetros para que o estado final fosse definido. 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.

3.1 Definição de um Autômato Finito Determinístico para as Classes Gramaticais

Houve a necessidade de um estudo através da verificação e comparação das classes gramaticais da Língua Portuguesa e as usadas e compreendidas pelos surdos ao se comunicarem ou ao redigirem algum texto, afim de que fosse fornecido ao autômato o estado inicial de cada um de seus símbolos de entrada e, conforme a LIBRAS, permitisse a transição que determinasse o estado final das classes gramaticais e das frases. Para as classes, em geral, tem-se a seguinte transição, representada na figura 2(a).



$$F: f(C_P) \rightarrow \gamma \text{ , onde } \gamma \in C_{s.}.$$

Figura 2- (a) Autômato Classe Gramatical (b) Função Transição

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 [6], é uma função parcial, pois é uma relação onde cada elemento do domínio está relacionado com, no máximo, um elemento do contradomínio, 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 efetua uma aplicação sucessiva da função programa sobre cada símbolo desta, até uma condição de parada aparecer. Existirá, portanto, uma linguagem aceita pelo autômato, a qual será o conjunto de todas as palavras pertencentes ao alfabeto, inclusive as vazias, que forem aceitas pela gramática. No caso, a linguagem é formada pelas classes gramaticais usadas, assim como suas variações, conforme a utilização dos surdos. Além disso, também existirão casos em que palavras não serão aceitas por esta gramática, sendo formado pelas classes gramaticais e suas variâncias não utilizadas e, assim, serão rejeitadas, formando um conjunto de palavras rejeitadas.

A função de transição tem a forma genérica apresentada na figura 2(b). Como uma ou mais funções poderão gerar resultados em γ estes, obrigatoriamente, deverão pertencer ao conjunto $C_{s.}$, isto é, os resultados gerados a partir desta função deverão pertencer ao conjunto formado pelas classes gramaticais usadas e compreendidas pelos usuários da LIBRAS, o que acarretará na modificação de algumas classes gramaticais, afim de que o texto final seja

aquele definido como compreensível pelos surdos. Por isso, para cada classe $C_{P,}$ poderá haver uma ou mais subfunções.

3.1.1 Artigos

Verificou-se o uso e compreensão desta classe gramatical pelos surdos, desde que antecedendo o substantivo. Logo, nos textos em que houver a ocorrência de artigo antes de outras classes gramaticais que não o substantivo, estes serão ignorados. A definição do autômato pode ser observada na figura 3(a) e 3(b).

3.1.2 Adjetivo

Por exercerem importante papel na transmissão do sentido das frases, esta classe continuará a ser utilizada no texto final após o processamento do autômato, pois funciona da mesma forma que na Língua Portuguesa. A definição do autômato pode ser observada na figura 3(c).

3.1.3 Numeral

Classe gramatical necessária pois, expressa ordem numérica, seguindo também a estrutura da Língua Portuguesa. A definição do autômato pode ser observada na figura 3(d).

3.1.4 Pronome

Ao analisar os textos escritos pelo grupo pesquisado, muitos destes, mostraram que pronomes do tipo pessoal oblíquo, como *lhe(s), vos, te* , e *ti* e os pronomes oblíquos reflexivos como *me* e *mim*, não são usados tão freqüentemente.

Assim, definiu-se um subconjunto de pronomes que permaneceram na tradução. Estes, foram denominados de *PronomesGlosa* e são: *eu, nós, ele (a), você(s), comigo,* e *contigo*; os demonstrativos como *esse (a), isso, este (a), isto, aquilo, aquele (a) e aquilo*; os possessivos mais usados como *meu, minha, teu, tua, seu e sua*; os interrogativos, como *quem, qual, quais e os quais,* quando usados no início das frases e na terceira pessoa do discurso como na Língua Portuguesa. A definição do autômato pode ser observada na figura 5(a)

3.1.5 Advérbios

Percebeu-se o uso somente de advérbios mais simples, tais como: *sim, talvez, certo, errado, muito, bastante, meio, todo, nada, demais, que, quando, quanto, bem, mal, quase, como, aqui, ali, aí, além, dentro, atrás, perto, longe, através, não, agora, hoje, depois, ontem, cedo* e *tarde*. Assim, para que o sentido das frases e textos no final da transição possuam valor, o autômato seguiu a modelagem apresentada na figura 5(b).

3.1.6 Preposição

Nesta classe, percebeu-se que ocorre o uso incorreto da preposição ou a omissão desta pelos surdos. Além disto, muitas vezes há a colocação de preposição desnecessária em seus textos. No entanto, há a necessidade do texto final possuir algumas das preposições mais usadas pelos surdos ao se comunicarem pela língua de sinais, como *"de"* e "*para"* e ainda as flexões "*da(s), do(s)"*. Logo, estas irão constituir o grupo denominado *PreposiçõesGlosa*, conforme visto na figura 5(c).

3.1.7 Conjunção

Nesta classe gramatical, percebeu-se que as mais utilizadas são "*mas", "e", "pois", "também", "ou", "antes", "que", "porque", "como"* e "*se"*, que formam o grupo de *ConjunçõesGlosa*. A definição do autômato pode ser observada na figura 5(d).

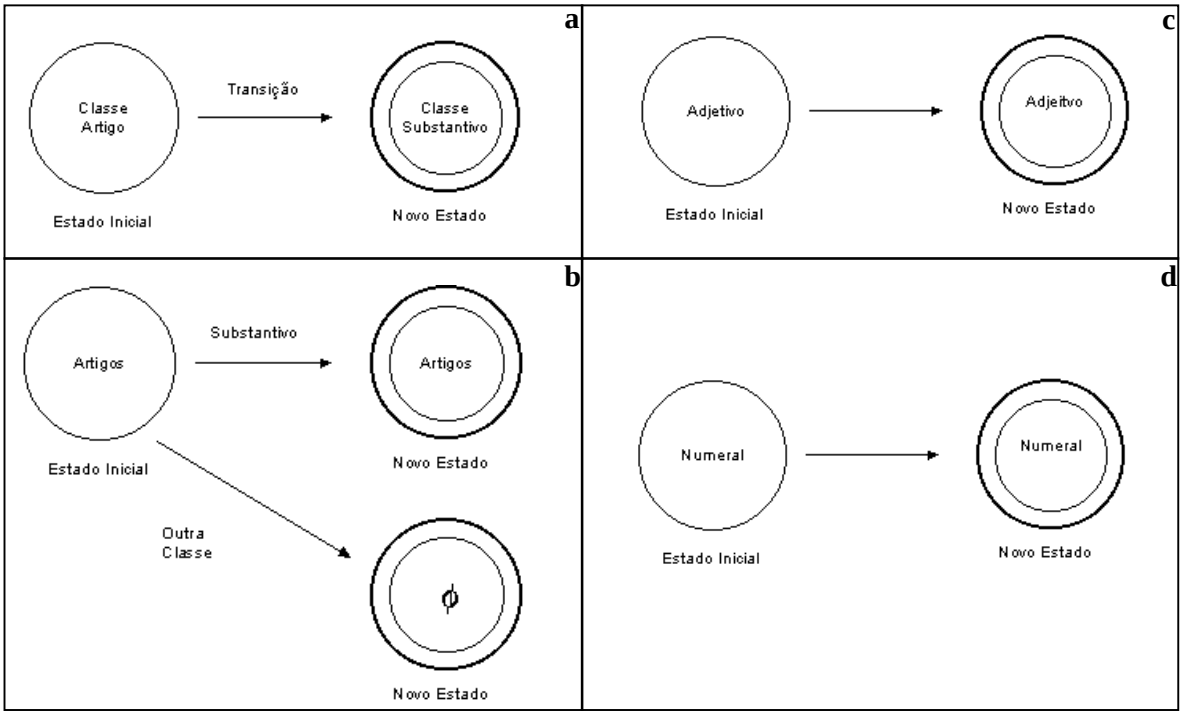


Figura 3- (a) Autômato Ocorrência Artigo (b) Autômato Artigo (c) Autômato Adjetivo (d) Autômato Numeral

3.1.8 Interjeição

Conforme o estudo realizado nos textos redigidos pelo grupo de surdos pesquisados, esta classe gramatical não é utilizada tão freqüentemente, ou seja, é mínimo. Portanto, ocorrerá a eliminação da mesma, conforme o autômato visualizado na figura 4(a).

3.1.9 Verbos

Verbos são um assunto delicado e, para melhor defini-los, estão sendo realizados mais alguns testes com os alunos da Escola de Ensino Médio Concórdia para Surdos de Santa Rosa, afim de que se possa encontrar a forma mais correta e que não prejudique a compreensão do texto nem se diferencie de sua cultura, a qual deve ser respeitada. Desta forma, foi modelado um autômato básico para os verbos, onde todos os tempos verbais dos textos serão passados para o infinitivo, conforme visto na figura 4(b), onde a função é definida por:

δ *(verbo)* $\rightarrow$ *função que transforma verbos dos textos em português para forma verbal no infinitivo.*

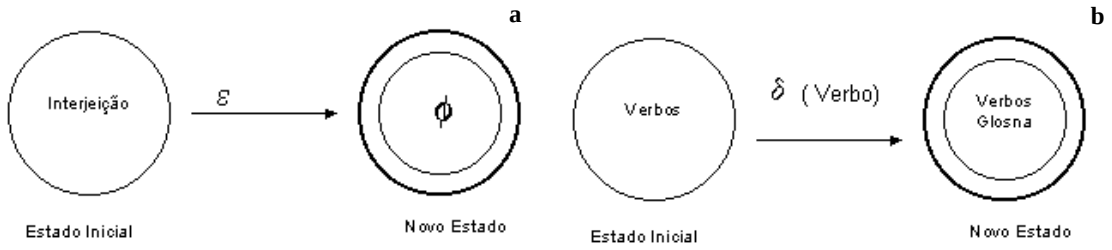


Figura 4 - (a) Autômato Interjeição (b) Autômato VerbosGlosa

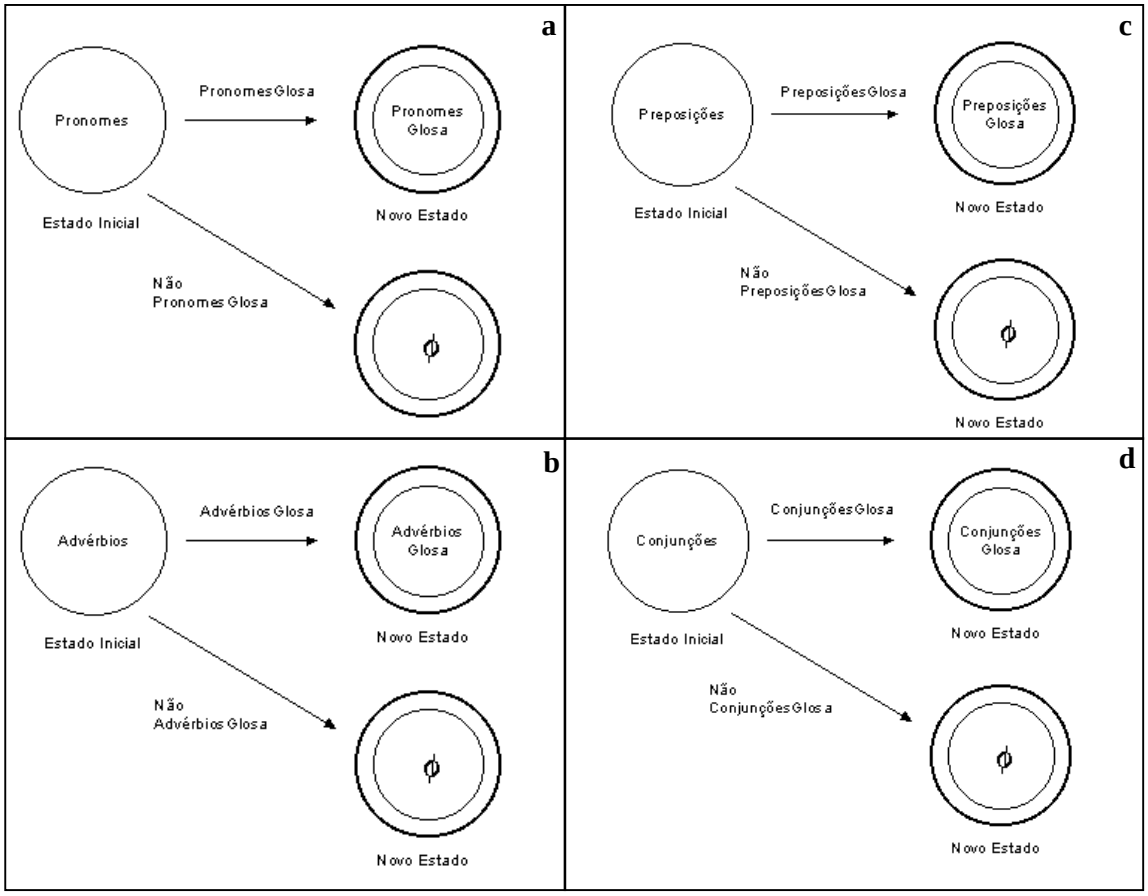


Figura 5 - (a) Autômato PronomesGlosa (b) Autômato AdvérbiosGlosa (c) Autômato PreposiçõesGlosa (d) Autômato ConjunçõesGlosa

3.2 Definição de um Autômato Finito Determinístico para o Texto

Em relação à ordenação das frases para os surdos, ainda estão sendo realizados testes que serão avaliados a fim de se encontrar a forma mais próxima da utilizada por eles. Afim de que o propósito desta modelagem seja concretizado, é necessário que a estrutura final dos textos seja compreensível e que não proporcione dúvidas, como acontece na maioria dos casos quando os surdos necessitam ler algum texto em português. Deve-se procurar uma forma em que as frases se estruturem sem ocorrer o risco de perder o verdadeiro sentido do texto, além de respeitar a língua–mãe destes. Assim, pode-se demonstrar a modelagem inicial para o autômato texto, que através dos testes que estão em andamento, nos proporcionarão organizá-lo conforme a sintaxe mais utilizada e compreendida pelos surdos ao se comunicarem.

4 Conclusões

Este trabalho procurou encontrar, através do estudo das Linguagens Formais, uma forma de solucionar o problema da não total compreensão de textos pelos surdos. Tal fato ocorre devido ao 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 [7], 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 pode-se 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.

Como a LIBRAS também é um língua como a Língua Portuguesa, a estruturação de frases, quando realizada manualmente, utiliza outros padrões. O mais significativo é o uso da topicalização, no qual os surdos partem de uma idéia genérica para a específica, fazendo com que os textos em português lidos por estes não sejam suficientemente claros, gerando incompreensão das informações.

A grande dificuldade encontrada foi a de determinar a ordem gramatical utilizada pela maioria dos surdos. Em relação à sintaxe, como citado em Pereira [7], esta tem tido atenção de estudiosos. No entanto, isto ocorre somente em relação à língua americana de sinais. Em relação á línguas de sinais usadas no Brasil, vários trabalhos estão sendo realizados nesta área da lingüística.

Devido a pouca compreensão de textos pelos surdos, ocasionada pela estrutura sintática das frases em português, foi procurado encontrar, através da modelagem de autômatos, os estados finais de cada classe gramatical encontrada nos textos eletrônicos para que fosse modelado o estado final do mesmo, tornando-o compreensível pela comunidade surda. Tal fato, no entanto, necessita ser complementado futuramente, devido à falta de um embasamento mais concreto em relação à estrutura sintática da glosa. Assim, pesquisas mais aprofundadas e testes estão sendo realizados, além dos já realizados anteriormente. Pretende-se, baseando-se nestes estudos, definir a estrutura frasal, de modo que posse ser realizada uma modelagem sintática dos textos e, conseqüentemente, sua implementação.

Devido as mudanças que estão sendo provocadas na sociedade com o avanço tecnológico, principalmente a popularização da *internet* e, conseqüentemente, a grande difusão de textos em formato eletrônico, o computador torna-se ferramenta essencial na popularização deste conhecimento. Desta forma, a definição desta ferramenta pode proporcionar à comunidade surda um recebimento maior das informações, fator básico 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 a esta comunidade o acompanhamento deste avanço.

5 Bibliografia

[1] QUADROS, Ronice Müller . 1997. **Educação de Surdos- A aquisição da Linguagem.** 1º Edição. Porto Alegre: Artes Médicas.

[2] BRITO, Lucinda Ferreira. 1995. **Por uma Gramática de Sinais.** Rio de Janeiro: Tempo Brasileiro.

[3] CEGALLA, Domingos Paschoal . 1996. **Novíssima Gramática da Língua Portuguesa.** 39ª Edição. São Paulo: Editora Nacional.

[4] VICENCIO, Aline Villa. 1994. **Uma Abordagem á Análise Sintática para Linguagem Natural.** Porto Alegre. Trabalho Individual de Pós-Graduação em Ciência da Computação da Universidade Federal do Rio Grande do Sul.

[5] 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.

[6] MENEZES, Paulo Fernando Blauth. 2000. **Linguagens Formais e Autômatos.** 3ª Edição. Porto Alegre: Editora Sagra - Luzzatto.

[7] PEREIRA, Maria Cristina da Cunha. **Aspectos Sintáticos da Língua Brasileira de Sinais.**1990.