

Baalbaki, A. Caldas, B (orgs) Instrumentos Linguísticos: usos e atualizações. Araruama: Cartolina Editora. 2014:155-188

BANCO DE DADOS E SISTEMAS DE TRANSCRIÇÃO PARA AS LÍNGUAS DE SINAIS

Tanya Amara Felipe¹

INTRODUÇÃO

A ciência da computação desenvolveu-se a tão ponto que gerou um mundo virtual que tem definido e controlado o mundo concreto através da internet e de inúmeros satélites que circulam em nossa órbita terrestre.

Pelo volume de páginas em várias línguas, disponibilizadas a cada segundo no ciberespaço, está sendo possível criar vários Bancos de Dados com corpora variados em várias línguas, que têm propiciado avanços nas pesquisas em Processamento de Linguagem Natural – PLN para a tradução automática ou semiautomática, a partir de modelos de *parsing* e de modelos estatísticos. Essas pesquisas são embasadas pela Linguística Computacional e pela Lexicografia Computacional que vêm trabalhando de modo interdisciplinar com pressupostos da Linguística, da Lexicografia e da Linguística de Corpus para a descrição e tradução de línguas naturais e criação de linguagens de computação.

Por isso, já é possível utilizar tradutores online que traduzem, com uma qualidade razoável, textos de várias línguas nas modalidades oral e escrita.

A partir da década de 90, além das pesquisas sobre línguas de modalidade oral-auditiva, começaram as pesquisas com as línguas de modalidade gestual-visual que estão propiciando o surgimento de Bancos de Dados com corpora específicos para as línguas de sinais.

Assim, nesse capítulo serão apresentados os sistemas que vêm sendo utilizados para o processo de tradução das línguas de sinais, quando também será relatado como se deu o processo de pesquisa para a criação de um protótipo de tradutor para a Libras.

¹ Professor Titular da Universidade de Pernambuco – UPE; Pesquisador do CNPq; membro do GP [Design de interação para a inclusão e o desenvolvimento social](#) -UFPR; Consultor da RNP; Pesquisador do CNPq; Proativa Soluções em Tecnologia/ PROATIVA – PE, Recife e Consultor da Fundação para Inovações Tecnológicas - FITEC-PE - GT PRODEAF. Site: www.librasemcontexto.org; tanyafelipe@gmail.com

1. AS LÍNGUAS DE SINAIS E SUAS ESPECIFICIDADES

Quando se trabalha com uma língua, às vezes, totalmente desconhecida e até ágrafa, como as pesquisas com línguas indígenas e línguas de sinais, é muito importante fazer uma coleta de dados e que esta seja feita com usuários que tenham bom domínio de sua língua para não haver problemas de interferência linguística do pesquisador ou para evitar a não correspondência dos dados com a realidade.

Em pesquisas com línguas que não têm uma tradição de estudos gramaticais, prioritariamente, está se buscando a regularidade, a gramática a partir do uso. Daí a necessidade da coleta dos dados.

Com relação às pesquisas sobre a Libras, depois da publicação de Gama (1875)², que é um glossário com explicações sobre os sinais desenhados da Língua de sinais Francesa, começaram com o livro de Oates (1969), que é um glossário: fotos de sinais, palavra correspondente em português e descrição para a sinalização, depois com o livro de Hoemann et al. (1983) que, dividido em 5 partes, aborda questões de fundamentação linguística, “aplicações educacionais e religiosas da Linguagem de sinais do Brasil, glosa em português e tradução dos modelos em vídeo-taipes e um léxico: linguagem de sinais do Brasil/português” com desenhos de sinais, Esses dois livros, até a década de noventa, eram utilizados por instrutores surdos para ensinaram a sua língua para ouvintes.

Mas as pesquisa sobre a Libras na área da linguística começaram a partir de 1982, a partir da dissertação de sobre a ordem sintática e a repetição na Libras (Namura, 1982) e sobre a estrutura temporal (Castro, 1982), seguida da dissertação sobre as categorias gramaticais na Libras, classificadores e sua estrutura frasal a partir de tipo de verbo (Felipe, 1988/89/91); a dissertação e artigos sobre pronomes (Berenz, 1989/93/96), (Berenz & Ferreira Brito, 1987); artigos os classificadores (Ferreira Brito, 1989), (Ferreira Brito et al., 1995); monografia e dissertação sobre narrativas pessoais (Felipe, 1991), (Caldas, 1992); artigo sobre a Negação (Ferreira Brito & Langevin, 1994); dissertação sobre a interação e construção do sistema gestual por crianças surdas (Pereira, 1989); dissertação sobre a aquisição do

² O INES reeditou esse livro. Ver em Felipe (1998 - Apêndice) a análise sobre esse livro.

parâmetro configuração de mão (Karnapp, 1994); dissertação sobre as Categorias Vazias Pronominais (Quadros, 1995) e dissertação e artigo sobre ato de fala (Faria, 1995), (Ferreira Brito, 1995).

Essa área de pesquisa sobre a Libras se consolidou em 1989, quando foi criado o GT Linguagem e Surdez, antigo GELES, foi aceito enquanto GT da ANPOLL – Associação Nacional de Pós-graduação em Letras e Linguística.

Atualmente, há inúmeras pesquisas sobre a Libras, material didático-pedagógico, dicionários e algumas instituições desenvolvendo pesquisa na área de tradução semiautomática e criação de avatar.³

Devido à tridimensionalidade das línguas de sinais e o não registro, da grande maioria delas, através de uma escrita, desde o início das pesquisas linguísticas sobre essas línguas, os pesquisadores têm criado e aperfeiçoado sistemas de transcrição para descrição de seus dados coletados através de filmadoras e armazenados em fitas, disquetes, CD, DVDs e computador.

Atualmente, com os avanços tecnológicos, tem surgido sistemas de marcação que tem permitido transcrever e analisar, através do computador, esses dados filmados previamente. Os registros podem ser feitos através de sistemas de transcrição ou de escrita de sinais. Esses corpora diversos podem ser armazenados em bancos de dados para utilização em sistemas de decodificação no processo de tradução.

Mesmo já existindo dois tipos de escrita e vários sistemas de transcrição para a representação escrita das línguas de sinais, para uma verdadeira acessibilidade dos usuários de uma língua de sinais, a preferência deles tem sido a criação de avatar que sinalize o texto escrito ou falado decodificado para a tradução.

Como nesse processo de tradução confrontam dois diferentes suportes – bidimensional, no caso da língua oral-auditiva, e tridimensional, no caso da língua gestual-visual – o trabalho de tradução implica uma subdivisão do processo em duas etapas muito distintas:

- (i) Tradução da língua oral-auditiva para algum tipo de transcrição simplificada da língua gestual-visual, cuja saída provida pelo tradutor dever obrigatoriamente respeitar as necessidades estabelecidas pelo sintetizador e

³ Para maior compreensão sobre AVATAR, consultar os seguintes endereços:

<http://pt.wikipedia.org/wiki/Avatar>

<http://embots.dfki.de/doc/Kippetal11.pdf>

http://www.ntid.rit.edu/sites/default/files/cat/NTID-SigningAvatar_20Mar2012_Final.pdf

http://www.academia.edu/3111707/Realistic_Avatars_for_Sign_Language_Synthesis_Poster_#

<http://www.youtube.com/watch?v=RarMKnjqzZU>

- (ii) Síntese de fala (no sentido lato), em que a transcrição é “executada”, ou seja, convertida num enunciado gestual-visual que é a parte gráfica relativa à síntese dos sinais, realizada pelo avatar.

Mas a criação desse avatar tem representado um grande desafio porque, para a sintetização da fala (em sentido lato), não é suficiente criar uma pessoa virtual gesticulando porque não se trata de uma animação gráfica simplesmente. Esse avatar tem que ser inteligente, ou seja, tem que sinalizar a partir da decodificação do nível fonológico de uma língua de sinais, como acontece com os sintetizadores de fala.

A partir de pesquisas para tradução automática ou semiautomática de línguas de sinais e de pesquisas para a produção de “avatar inteligente”, nesse capítulo serão apresentados e analisados os sistemas de transcrição utilizados para a criação de Bancos de Dados e de registro de língua de sinais, refletindo como todas essas pesquisas para a descrição linguística de decodificação têm propiciado avanços na elaboração de dicionários bilíngues e de estudo comparativo de regras no processo de tradução automática e elaboração de avatar que estão sendo elaborados para traduzir uma língua de modalidade oral auditiva para uma língua de modalidade gestual-visual.

2. SISTEMAS DE TRANSCRIÇÃO PARA AS LÍNGUAS DE SINAIS: DE BÉBIAN AO HAMNOSYS

Embora desde o século XVII, a partir das primeiras escolas para surdos na Espanha e França⁴, as línguas de sinais desses países tenham servido de base para propostas metodológicas para a educação de surdos, a proposta do Abade l'Épée, conhecida como método misto, foi a mais difundida. No entanto, as línguas de sinais adquiriram o status de língua de fato somente após as pesquisas de Stokoe (1960) e seu grupo de pesquisa que descreveram a *American Sign Language* - ASL, a partir de pressupostos teóricos da linguística estruturalista, apresentando uma proposta de transcrição em que era possível apresentar os sinais de uma língua de modalidade gestual-visual em seu nível fonológico.

Segundo, Oviedo (2009:13), Bébian (1825) precedeu Stokoe nessa pesquisa de descrição linguística de uma língua de sinais, uma vez que Stokoe, sob o mesmo critério de Bébian e conhecedor de seu trabalho pioneiro, trabalhou também com as informações sobre a configuração de mão e sua orientação espacial e o sistema de transcrição desenvolvido por

⁴ Consultar o acervo do Instituto nacional de Educação de Surdos – INES, Rocha (2008) e Ballesteros e Villabrille (1863), edição 2010 – Editorial Maxtor. Valladolid. Espanha.

Bébian para descrever a língua de sinais francesa - LSF, mas o mérito de Stokoe foi ter refinado a descrição destas categorias em dois parâmetros independentes do sinal - Tabulação: localização de um sinal e Designador: forma da mão – e ter acrescentado mais uma categoria - Signation: tipo de movimento articulado, além de ter criado um sistema de transcrição diferente do de Bébian. Fischer (1995) também analisou esse sistema de Bébian.

Tendo sido Bébian (1817) professor em uma escola de surdos e não querendo trabalhar com o método misto de Abade L'Épée, ele foi também pioneiro ao optar uma educação bilíngue em que a língua de sinais dos alunos surdos fosse também ensinada e aprendida pelos alunos e professores, além de ter proposto que a LSF se tornasse a língua de instrução da escola. No entanto, sua contribuição tanto para os estudos linguísticos das línguas de sinais como para a educação de surdos não teve uma repercussão e reconhecimento. Por isso, somente a partir de Stokoe, outros pesquisadores iniciaram suas pesquisas, acrescentando novos parâmetros e novos detalhamentos ao seu sistema de transcrição (Klima & Bellugi, 1979; Battison, 1978), sendo possível perceber a abrangência da influência de seu trabalho a partir da árvore genealógica dos sistemas de transcrição existentes, proposto por Miller (1994).⁵

Um desses sistemas é o *Hamburg Sign Language Notation System* – HamNoSys, desenvolvido na Universidade de Hamburgo (Prillwitz, Leven, Zienert, Hanke, Henning, 1989)⁶. Ele é um sistema unidimensional linear semelhante aos sistemas alfabéticos escritos, uma vez que utiliza uma ordenação arbitrária linear de símbolos para os parâmetros: configuração de mão, localização e movimento e tem por finalidade descrever linguisticamente sinais isolados. Esse sistema, baseado no sistema de notação de Stokoe, foi ampliado para um conjunto de aproximadamente 200 símbolos que permitem uma transcrição fonética de qualquer característica potencialmente significativa de qualquer sinal em qualquer língua de sinais e, por isso, tem sido o sistema de transcrição o mais utilizado atualmente.

3. SISTEMAS DE ESCRITA PARA LÍNGUAS DE SINAIS

Em 1974, surgiu a primeira proposta para que as línguas de sinais deixassem de ser “ágrafas” e pudessem ter registro escrito formal para transmissão através de uma escrita. Dessa forma, os surdos poderiam ler textos em suas línguas de sinais e produzir textos informativos, acadêmicas, literários, entre outros, como também as crianças surdas poderiam ser “alfabetizadas” em sua primeira língua (Felipe, 2012).

⁵ Ver Amaral (2012) que apresenta vários sistemas de transcrição em sua tese.

⁶ <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html>

Atualmente, segundo os relatórios sobre a listserv *SignWriting* (sw-l@majordomo.valenciacc.edu), parece haver, pelo menos, 14 escolas no mundo que estão usando esse sistema e, atualmente, é a forma de escrita de 27 línguas de sinais. Aqui no Brasil, esse sistema de escrita foi introduzido, em 1996, através do professor da PRC-RS, Antônio Carlos da Rocha Costa. Sua aluna surda, Marianne Stumpf, passou a estudar, divulgar e ensinar esse sistema no Ensino Básico e, atualmente, como Prof. Dr. Da UFSC, é responsável pela disciplina *Signwritng* no Curso de Graduação em Letras Libras.⁷

Atualmente, há várias publicações em *signwriting*⁸, e algumas experiências de sua utilização em sala de aula no Ensino Básico, sendo a Escola Especial Concórdia, em Porto Alegre, a pioneira nesse trabalho realizado por sua professora Marianne Stumpf. No Ensino superior, esse sistema de escrita se constituindo enquanto disciplina curricular nos cursos de graduação em Letras/Libras existentes em várias cidades brasileiras.

Hulst (1993, 2000) apresentou um estudo comparativo detalhado sobre os sistemas de escrita *SignWriting* de Sutton (1974) e o ASL-Phabet de Supalla & Blackburn (2003) com relação ao o sistema de transcrição HamNoSys⁹, analisando suas diferenças com relação às finalidades e potencialidades.

Corradi (2007), em sua dissertação, faz uma análise exploratória e descritiva de ambientes digitais, apresentando um estudo minucioso de vários softwares para o *signwriting*, tais como: SignText (criação de documentos); SignPuddle (criação de dicionários); SignBank (criação de bases de dados); IMWA (Alfabeto em *SignWriting* em movimento); o SWML (*SignWriting Markup Lang*); o *SIGN WEBMESSAGE*; o SWEDIT e o AGA-Sign (Denardi et al., 2005), reforçando a abrangência e possibilidades de usos desse tipo de escrita para as línguas de sinais.

O *signwriting* e o ASL-Phabet têm a mesma finalidade enquanto propostas para escrita de língua de sinais, mas sem entrar em detalhes com relação a questões pedagógicas e culturais no que tange à necessidade de uma escrita para as línguas de sinais e sem discutir também sobre as vantagens e desvantagens desses dois sistemas de escrita, por serem sistemas que exigem conhecimentos específicos para sua utilização e compreensão e necessitarem também de adaptações para a utilização em PNL para a tradução automática ou semiautomática, estes

⁷ Conferir em <http://paulohenriquelibras.blogspot.com.br/2011/07/grandes-nomes-grandes-contribuicoes-02.html>

⁸ Conferir em: <http://www.signwriting.org/brazil/>; http://www.libras.ufsc.br/colecaoLetrasLibras/eixoFormacaoEspecifico/literaturaVisual/assets/369/Literatura_Su_rda_Texto-Base.pdf; Escola Estadual de Educação Especial: <http://www.signwriting.org/brazil/brazil30.html>

⁹ Supalla, Sam & Blackburn, Laura (2003) <http://en.wikipedia.org/wiki/ASL-phabet> . <http://aslphabet.com/home.html>

sistemas não vêm sendo utilizados para transcrição de dados pela maioria dos pesquisadores que estão preferindo trabalhar com sistemas de transcrição em glosa.

4. SISTEMA DE TRANSCRIÇÃO EM GLOSA

O sistema de transcrição de Stokoe foi criado para descrição fonológica de sinais isolados e, além de exigir um estudo para sua compreensão, ele possui várias lacunas com relação às representações para marcas não-manuais (Felipe, 2013) e marcas morfossintático-discursivas, por isso, os linguistas que começaram a descrever as línguas de sinais, sentiram a necessidade de criar um sistema mais acessível que também desse conta de descrever esses aspectos morfossintático-discursivos de uma língua de modalidade gestual-visual, que não possuía uma escrita, e que fosse compreensível para os leitores que não conheciam uma língua de sinais. Assim, surgiram sistemas de transcrição que, utilizando também sinais gráficos e tipográficos podiam representar especificidades morfossintático-discursivas da língua de sinal transcrita em glosa, entendendo esta como uma representação semântica de um sinal, através de uma unidade fonográfica escrita de uma língua oral auditiva.

Aqui no Brasil, a partir desses sistemas de transcrição que vinham sendo utilizados por linguistas em outros países, para a descrição morfossintático-discursiva da Libras foram utilizadas essas convenções, mas foi preciso fazer adaptações para representar a Libras a partir de glosas em língua portuguesa.

Felipe (1988; 1989), para apresentar seus dados sobre a estrutura frasal, coletados na Associação de Surdos de Pernambuco, através de filmagens de conversações e narrativas espontâneas, criou um sistema de transcrição em glosa que representava sinais da Libras através de palavras em português e de sinais gráficos e tipográficos e, por isso, teve que fazer modificações e acréscimos às convenções utilizadas por linguistas que utilizavam glosas em língua inglesa.

Para as suas monografias de conclusão de disciplinas no curso de doutorado, de 1991 a 1993, suas publicações e sua tese de doutorado em que coletou dados para seu estudo das relações sintático-semântica de 798 verbos da Libras (Felipe, 1998 – volume 2), foi necessário ampliar esse sistema para transcrever esses dados coletados - exemplificações de usos desses verbos filmados - em sete fitas de vídeo cassete que corresponderam a 1.332 páginas transcritas, armazenadas em 3 disquetes. Além das filmagens foram coletados também fotos e desenhos dos sinais da maioria dos verbos pesquisados, em que foram utilizados sinais gráficos (setas e outros recursos gráficos) para inclusão de informações relacionadas ao movimento e direcionalidade. Esse sistema de transcrição em glosa, denominado de sistema

de transcrição em palavra, foi publicado pela primeira vez nos anais do II Congresso da Assel-Rio (Felipe, 1993: 216-23), assim apresentado:

4.1. Metodologia para descrição em Língua de sinais:

O Corpus: os dados analisados foram retirados de filmagens de narrações pessoais de surdos, Felipe (1991), de filmagens para licitação de dados, quando se pedia a informantes surdos que fizessem frases com EVENTO selecionado previamente, e de um levantamento de tipos de EVENTO da LSCB feito com surdos, professores de LSCB.

Sistema de Notação: Para a transcrição dos dados da Língua dos Sinais Americana (ASL), há dois sistemas de notação. Um usado por Friedman (1976), Liddell (1978), Klima & Bellugi (1979) e Padden (1983), entre outros; e um usado por Kegl (1985).

O primeiro, chamado de "**Sistema de notação em palavras**", utiliza palavras da Língua Inglesa para representar os itens lexicais da ASL correspondentes e uma simbologia específica para os traços não-manuais que são feitos simultaneamente com esses itens, como: marcadores de aspecto, frases interrogativas, negativas, etc. Este sistema também foi usado por Deuchar (1984) para transcrever a Língua dos Sinais Britânica (BSL).

O segundo, proposto por Kegl, é o "**Sistema de notação de base Locacional**". Este tipo de notação mostra a estrutura morfo-sintática da ASL, podendo-se apreender, por exemplo, os itens que se assemelham ou se diferenciam devido ao tipo de raiz.

Para a transcrição dos dados da Língua dos Sinais dos Centros Urbanos Brasileiros (LSCB), foram utilizadas as convenções já conhecidas e foram criadas outras para pesquisas anteriores, Felipe (1988,1989) e para esta:

1. Os itens lexicais da LSCB, para efeito de simplificação, foram representados por itens lexicais da Língua Portuguesa (LP) em letras maiúsculas. Quando era necessário mais de um item da LP para representar um item da LSCB, estes foram escritos justapostos, diferenciando das palavras compostas em LSCB, que foram representadas por palavras da LP separadas por hífen. Exemplos:

- CASA "casa", HOMEM "homem", ESTUDAR "estudar", LEÃO "leão";
- CORTARFACA "cortar com faca", CORTARTESOURA "cortar com tesoura", NÃOQUERER "não querer", COMERMAÇÃ "comer maçã";

- CASA-ESTUDAR "escola", MULHER-BENÇÃO "mãe", HOMEM-BENÇÃO "pai", COMER-MEIODIA "almoçar";

2. Quando não há em LSCB um sinal para um conceito já codificado em língua portuguesa, por empréstimo, esse item lexical é expresso pelo alfabeto manual (datilologia). Este recurso foi representado pelas letras do alfabeto da LP separadas por hífen. Exemplos:

- J-O-Ã-O "joão", A-N-E-S-T-E-S-I-A "anestesia";

3. Os traços não-manuais: expressões faciais e/ou expressões corporais, que coarticulam-se com os itens lexicais, foram representados por uma linha pontilhada acima do item ou frase, intercalada por outro símbolo representando o tipo de traço. Exemplos:

a) Tipo de frase:

...i... (interrogativa),
...n... (negativa),
...e... (exclamativa),
...imp... (imperativa),
...i/e... (interrogativa-exclamativa),

b) marcadores do discursivo:

...t... (topicalização),
...en... (ênfase),
...mr... (mudança de referente),
...mt... (mudança de turno),

c) marca de aspecto:

...cont... (aspecto continuativo),
...iter... (iterativo),
...dur... (durativo),

d) como o advérbio de modo pode ser uma expressão facial e/ou corporal coarticulada com o item, foi representado o tipo de expressão acima do item com seu sentido:

- exp.f "espantado" (a expressão facial mostra espanto),
- exp.f/c "indiferentemente" (as expressões facial e corporal mostram indiferença);

4. Quando há um bimodalismo, ou seja, um item da LSCB é coarticulado com um de LP, este som foi representado pela transcrição acima do item da LSCB. Exemplos:

<vou> <viaja>

IR, VIAJAR;

5. Em LSCB, as categorias semânticas COISA e PROPRIEDADE, morfologicamente, não têm desinência para gênero e número. Para COISA, os gêneros são formados, sintaticamente, pela posposição dos itens lexicais HOMEM e MULHER, respectivamente, somente quando se deseja explicitá-los. Mas, geralmente, as duas categorias mencionadas ficam na forma neutra, como os **advérbios**, **intensificadores** e **pronomes indefinidos**. Esta **forma neutra** foi representada pelo símbolo "@", justaposto ao radical do item. Exemplos:

- AMIG@, FRI@, MUIT@, TOD@;

6. As categorias semânticas IDENTIDADE, POSSE e LUGAR, consideradas enquanto categorias gramaticais como sendo os pronomes pessoais, os possessivos, os demonstrativos e os advérbios de lugar são dêiticos em LSCB e se configuram, espacialmente, em relação ao emissor do ato de fala, por isso foram representados com relação às três pessoas do discurso:

- **IDENTIDADE**: ind_{1s} "eu", ind_{2s} "você", ind_{3s} "ele/ela"; ind_{1d} "nós dois/duas", ind_{2d} "vocês dois/duas"; ind_{3d} "eles dois/elas duas"; ind_{1p} "nós", ind_{2p} "vocês", ind_{3p} "eles/elas";

onde: ind = indica;

1s, 2s, 3s = 1ª 2ª e 3ª, pessoas do singular;

1d, 2d, 3d = 1ª 2ª e 3ª pessoas do dual;

1p, 2p, 3p = 1ª 2ª e 3ª pessoas do plural;

- **LUGAR**: loc_i "este,esta,isso/aqui"; loc_j "esse, essa, isso/aí"; loc_k "aquele,aquela, aquilo/alí";
onde: loc = locativo;

variáveis: i = ponto próximo à 1ª pessoa,

j = ponto próximo à 2ª pessoa,

k, k' = pontos próximos às 3ªs pessoas;

- **POSSE**: ind_i "meu/minha/ nosso"; ind_j "teu(s)/tua(s); ind_k "seu/sua/dele/dela"; ind_{k'} "seu/sua/dele/dela. O "k'" foi usado quando mais de um ponto estava sendo convencionado para 3ªs pessoas diferentes.

7. Quando a categoria EVENTO trazia também a informação sobre a direção ou número-pessoal, foi representada da seguinte forma:

- **EVENTO de raiz de...** -> $i^{---}K: i^iR_K, i^iVIAJAR_K;$

- **EVENTO de raiz ...para** -> $k^{---}i: k^kVIR_i, k^kABRIR-GAVETA_i;$

- **EVENTO com flexão número-pessoal:** em subscrito, foi identificada, à direita, a pessoa do discurso que é o sujeito- origem e, à esquerda, a pessoa que é o objeto-meta da frase onde o verbo foi coletado:

1_sDAR_{2s} "eu dou para "você",

$2_sPERGUNTAR_{3p}$ "você pergunta para eles/elas";

- **EVENTO multidirecional** -> em subscrito, foi identificada, à direita, a origem e, à esquerda, a meta do movimento, como também seu sentido:

$k_dANDAR_{k'e}$ "andar da direita (d) para à esquerda (e) a partir dos pontos convencionados para o movimento de uma 3ª pessoa".

Em trabalho posterior, Felipe (1997, 1998a, 1998b), fez algumas modificações e acréscimos nesse seu sistema de transcrição:

1. um sinal, que é traduzido por duas ou mais palavras em língua portuguesa, será representado pelas palavras correspondentes separadas por hífen. Exemplos: CORTAR-COM-FACA, QUERER-NÃO "não querer", MEIO-DIA, AINDA-NÃO, etc;
2. um sinal composto, formado por dois ou mais sinais, que será representado por duas ou mais palavras, mas com a idéia de uma única coisa, serão separados pelo símbolo ^ . Exemplos: CAVALO^LISTRA "zebra";
3. o sinal soletrado, ou seja, uma palavra da língua portuguesa que, por empréstimo, passou a pertencer à LIBRAS por ser expressa pelo alfabeto manual com uma incorporação de movimento próprio desta língua, está sendo representado pela datilologia do sinal em itálico. Exemplos: *R-S* "reais", *A-C-H-O*, *QUM* "quem", *N-U-N-C-A*, etc;
4. Os traços não-manuais: expressões facial e corporal, que são feitos simultaneamente com um sinal, estão representados acima do sinal ao qual está acrescentando alguma idéia, que pode ser em relação ao:

a) tipo de frase: $\overset{\text{interrogativa}}{\text{ou... i ...}} \overset{\text{negativa}}{\text{ou ... neg ...}} \text{etc}$

Para simplificação, serão utilizados, para a representação de frases nas formas exclamativas e interrogativas, os sinais de pontuação utilizados na escrita das línguas orais-auditivas, ou seja: !, ? e ?!

b) advérbio de modo ou um intensificador: $\text{muito rapidamente exp.f "espantado" etc;}$

Exemplos: $\overset{\text{interrogativa}}{NOME} \overset{\text{exclamativo}}{ADMIRAR} \overset{\text{muito}}{LONGE};$

entanto, muitos professores do Ensino Básico que trabalham com alunos surdos, por desconhecerem que um sistema de transcrição não é uma escrita de sinais e nem uma forma errada “de surdos” escreverem português, têm desaconselhado a utilização de sistemas de transcrição, cuja finalidade não é nem escrever em português e nem é uma escrita de sinais, mas uma convenção para apresentar linguisticamente uma língua de modalidade gestual-visual para pessoas que não estão familiarizadas com a escrita da Libras.

5. SISTEMAS DE MARCAÇÃO

Muitos pesquisadores que trabalham com dados originais sentem dificuldades na busca de seus dados armazenados cassetes ou outras fontes multimídia porque nem sempre é fácil o seu manuseio, além das dificuldades para se efetuar estudos comparativos entre partes específicas em uma gravação. Mas a partir de novos softwares, já é possível a filmagem e reprodução de vídeo pelo computador, como também da conversão de modalidade oral para a modalidade escrita ou da escrita para voz; daí, têm surgido sistemas de marcação para o tratamento computacional e de notação de dados para diversas línguas, como também para as línguas de sinais.

Segundo Hulst (1993, 2000), para as línguas de sinais, já existe o *SIGNSTREAM*¹¹ e o ELAN¹² que poderiam ser considerados homólogos ao PRAAT¹³, que é um software para anotação do material de áudio. Esses sistemas permitem aos pesquisadores marcar os dados em vídeo, alinhando informações linguísticas com os dados originais, organizadas em várias linhas ou trilhas definidas pelo programa ou pelo usuário do programa, que pode fazer sua escolha com relação às propriedades linguísticas relevantes que quer analisar.

O *SignStream* é destinado principalmente à análise sintática de sinais, embora permita a inserção de algum detalhe fonético. No entanto, ele não vem sendo muito utilizado por estar disponível apenas para sistemas Mac. Por outro lado, embora o ELAN esteja focado mais para

¹¹ O SIGNSTREAM foi desenvolvido por pesquisadores da Universidade de Boston (<http://www.bu.edu/asllrp/SignStream/>) Neidle et al. (2002) American Sign Language Linguistic Research Project. Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Ashwin Thangal, Haijing Wang and Quan Yuan (2002). Large Lexicon Project: American Sign Language Video Corpus and SignLanguage Indexing/Retrieval Algorithm Vassilis Athitsos https://www.academia.edu/2845117/Large_Lexicon_Project_American_Sign_Language_Video_Corpus_and_Sign_Language_Indexing_Retrieval_Algorithms#

¹² O EUDICO Language Annotator foi (<http://www.mpi.nl/tools/>) foi criado pelo pesquisadores do Instituto Max Planck de Psicolinguística e está disponível através de seu site: <http://www.mpi.nl/tools/elan.html>

¹³ O PRATT foi desenvolvido por Paul Boersma e David Weenink (<http://www.fon.hum.uva.nl/praat/>).

a descrição fonética, ele permite a descrição de qualquer tipo de detalhe, além de ter a vantagem de está disponível para sistemas Mac, Windows e Linux.

O problema destes anotadores é o fato de terem sido concebidos para o trabalho com material de multimídia apenas, não sendo ainda possível a inclusão de fotografias, desenhos ou descrições de dicionário, mas mesmo com estas restrições, estes sistemas de marcação têm permitido a transcrição dos dados de línguas de modalidade oral-auditiva e gestual-visual coletados armazenados em multimídia e em cartão de memória.

O ELAN vêm propiciado a organização de bancos de dados com corpora específicos em várias línguas de sinais que poderão futuramente ser utilização no processo de tradução automática para várias línguas de sinais. Aqui no Brasil, grupo de pesquisa na USP, McCleary et al. (2010) e na UFSC, Christmann et al. (2014), estão organizado seus bancos de dados, utilizado esse sistema de marcação, além de outros iniciativas individuais para análise de dados sobre a Libras em dissertações e teses, com as de Nascimento (2013), Rodrigues (2014), entre outras.

6. SISTEMAS DE CODIFICAÇÃO

Todos esses sistemas descritos acima têm despertado a atenção de pesquisadores da área de computação que estão, juntamente com pesquisadores de outras áreas, criando ferramentas para permitir transcrição e descrição das línguas de sinais e o acesso a essas línguas pelos seus usuários no universo da informática. Esses avanços estão exigindo a criação de sistemas de codificação que são ferramentas de pesquisa que permitem que linguistas e pesquisadores de PNL possam formalizar descrições linguísticas para a classificação computadorizada de línguas de sinais para várias finalidades. Esses sistemas exigem alta explicitação, simplicidade, transparência e facilidade de armazenamento e recuperação de dados.

Segundo Hulst (1993, 2000), o mais extenso banco de dados focado em codificação para língua de sinais é o projeto SignPhon com códigos alfanuméricos condensados e detalhados para mais de 3000 sinais de Língua de Sinais Holandesa para outras línguas de sinais Crasborn, van der Hulst e van der Kooij (2001). Outro sistema é o SignTyp, que vem sendo considerado o sucessor do SignPhon . Esse novo sistema de codificação de sinais usa um vocabulário Inglês fechado para descrever sinais em tantos detalhes quanto o analista necessita e permite pesquisas simples e tipos de dados atômicos. Possuindo uma estrutura de banco de dados flexível e relacional, permite que os pesquisadores tenham mais liberdade

para fazer mudanças na estrutura de dados sem envolver o banco de dados ou administrador do site.

Outros sistemas de decodificação para a Libras têm sido testados ou criados aqui no Brasil, como por exemplo: XML2 (Antunes, 2011); F-Libras (Fabrício, 2007) e LIST (FENEIS-NILC, 2003). Este último será detalhado a seguir.

7. PROJETO TRADUTOR PORTUGUÊS-LIBRAS DA FENEIS

Em 2003, a Federação Nacional de Educação de Surdos – Feneis, juntamente com a equipe do Núcleo Interinstitucional de Linguística Computacional – NILC da Universidade Federal de São Carlos e da equipe da AcessibilidadeBrasil executaram o Projeto Tradutor Libras x Português.¹⁴

O objetivo dessa pesquisa foi desenvolver um sistema de tradução automática auxiliada por humanos (*human-aided machine translation*) que foi denominado PULØ – Portuguese-UNL-LIST deOralizer. O resultado final foi uma versão experimental de um sistema de tradução automática unidirecional de uma língua oral-auditiva, o português, para a representação linear (*Libras Script for Translation – LIST*) de uma língua gestual-visual, a língua brasileira de sinais – Libras, além de um Banco de dados com 8.000 sinais – Dicionário da Libras, de um Banco de dados com Filmagens e transcrição de histórias em quadrinho e de páginas de livros didáticos.

O PULØ, protótipo projetado pelo NILC para compor um dos módulos do sistema de tradução português-Libras, foi proposto como um sistema de tradução automática a) baseado exclusivamente em conhecimento linguístico e b) que utilizou a estratégia de tradução indireta por interlíngua. Para a sua concepção, foram desenvolvidos os seguintes recursos linguísticos: a notação LIST, o formalismo gramatical NL-UNL, a gramática português-UNL, o dicionário português-UNL, a gramática UNL-LIST e o dicionário UNLLIST. Além desses recursos, foram produzidos subsistemas computacionais projetados a partir dos aplicativos fornecidos pela *Universal Networking Digital Language Foundation*, de Genebra.¹⁵, detalhados por Martins, Pelizzoni e Hasegawa (2005).

¹⁴ Projeto financiado pelo MEC-SEESP/FNDE, cuja coordenação geral de pesquisa ficou sob a responsabilidade da Prof. Dr. Tanya Felipe, também coordenadora do GT-FENEIS, e a coordenação geral administrativa, sob a responsabilidade de Guilherme Lira, que também ficou responsável pela contratação da equipe de computação gráfica, - GT-AcessibilidadeBrasil, que iria construir o avatar. Esta parte do texto foi elaborada a partir do relatório da GT-NILC: Maria das Graças Volpe Nunes (coordenadora), Jorge Marques Pelizzoni, Juliana Galvani Gregghi, Ricardo Hasegawa, Ronaldo Teixeira Martins, juntamente com o relatório do GT-Feneis.

¹⁵ Para maiores informações sobre a *Universal Networking Language - UNL*, ver a webpage da UNDL Foundation, em <http://www.undl.org>.

Devido à falta de sincronia entre os desenvolvedores dos dois módulos do sistema: o GT-NILC e o GT AcessibilidadeBrasil -equipe da computação gráfica, o GT-NILC acabou por definir unilateralmente um formato de saída, – batizado de LIST, baseado na notação linear proposta por Felipe (1998), como será apresentado mais à frente.

7.1. LIST - Libras *Script for Translation*

LIST é uma proposta de transcrição de Libras utilizada computacionalmente, concebida especialmente para fins de tradução (semi)automática intermodal e síntese de fala na língua brasileira de sinais. Esse sistema de decodificação constituiu o formato de saída de PULØ e o formato de entrada do sintetizador de fala – segundo módulo que seria produzido pelo GT AcessibilidadeBrasil. Seu objetivo foi representar, de forma não ambígua, todas e apenas as informações necessárias para a geração de sinais da Libras a partir de uma sentença em língua portuguesa, considerando:

- a) que português e Libras são línguas distintas, não apenas em relação à estrutura (gramática e léxico), mas ao próprio suporte;
- b) que a sentença em língua portuguesa possui redundâncias (como a concordância de número, de gênero e de pessoa gramatical para todos os verbos) que foram suprimidas no processo de representação na Libras;
- c) que a sentença em língua portuguesa possui informações (como a ordem dos constituintes) que nem sempre são relevantes para a representação em Libras;
- d) que a sentença em língua portuguesa possui lacunas (informação sobre classificadores, elipses, anáforas e outras pró-formas, nominais e verbais) que deveriam ser preenchidas, ainda que de forma arbitrária, para que se pudesse assegurar a gramaticalidade e a semanticidade dos enunciados da Libras.

O LIST foi planejado para ter três partes distintas e interligadas: a) o vocabulário LIST; b) a estrutura frasal de LIST; c) a estrutura dos documentos em LIST. Nesse trabalho, será apresentado apenas o primeiro.

7.1.1. Vocabulário LIST

O vocabulário LIST foi constituído de *LIST Words (LWs)*, que correspondeu à representação linear (bidimensional) dos sinais da Libras, definido a partir do conjunto de itens lexicais de língua portuguesa. Cada LW foi composta de uma *headword*, necessariamente, seguida de uma matriz atributo-valor, opcionalmente.

Para as *headwords*, foi adotada a seguinte convenção: todos os sinais de natureza fonográfica (ou seja, que procuram imitar, em sua formação, a estrutura fonológica das

palavras da língua portuguesa, por meio da datilologia ou sinal soletado, foram representados em letras maiúsculas, sem hífen; os demais sinais, de natureza não-fonográfica (ou seja, logográfica, ideográfica, pictográfica, etc) foram representados por letras minúsculas¹⁶. A par desta convenção, foram adotados ainda os seguintes princípios:

- a) lexias simples da Libras que não selecionam classificadores serão representadas pelos itens lexicais da língua portuguesa correspondentes: ‘casa’, ‘estudar’, ‘criança’;
- b) lexias compostas e lexias complexas do português que correspondem a lexias simples da Libras serão representadas pelos itens lexicais da língua portuguesa correspondentes conjugados com hífen: ‘cortar-com-faca’, ‘querer-não’, ‘meio-dia’;
- c) lexias simples do português que correspondem a lexias compostas da Libras serão representadas pela combinação dos itens lexicais da língua portuguesa correspondentes conjugados pelo símbolo "^": ‘cavalo^listra’ (zebra).

Dada a estrutura morfológica dos sinais da Libras, as headwords de LIST não compreenderam nenhum sufixo flexional específico, sendo representados pelas formas do masculino singular, no caso dos substantivos e palavras de natureza substantiva; do masculino singular, no caso dos adjetivos e palavras de natureza adjetiva; e do infinitivo impessoal, no caso dos verbos. Espaços em branco foram proibidos dentro de headwords.

De resto, o espaçamento entre quaisquer elementos ficou livre, incluindo a possibilidade de inserção de quebras de linha na codificação de uma mesma lista.

As lexias simples da Libras que selecionam classificadores foram representadas pelas headwords seguidas de uma matriz atributo-valor. Esta seria uma lista, entre chaves, de atribuições do tipo “Atributo:Valor”, separadas por espaço em branco, que representou a indicação explícita da função sintática ou do caso semântico subcategorizado. Por exemplo, a LW correspondente ao item lexical do português “colocar”, que, em Libras, é um verbo classificador que concorda com o objeto a ser colocado e faz referência ao sinal do local onde se coloca o objeto, seria assim representada: ‘colocar{genobj:coisa-redonda loc:embaixo{refi}}’, indicando, neste caso, que a ação descrita pelo verbo e que foi recuperada na geração do sinal corresponde ao ato de colocar um objeto de forma arredondada embaixo de outro [i] referido anteriormente, abaixo do qual o novo sinal deveria ser gerado.

¹⁶ A opção pelas letras maiúsculas e minúsculas diverge da estabelecida em Felipe 2001 para evitar ambigüidade. Na proposta da autora, o item lexical (hipotético) "X-Y-Z" poderia corresponder, do ponto de vista computacional, (a) à lexia composta "X-Y-Z", que poderia ser representada por um único sinal (como "CORTAR-COM-FACA"; "MEIO-DIA"); e (b) à lexia simples "X-Y-Z", a ser representada por meio da datilologia (como "J-O-Ã-O").

O quadro abaixo ilustra a correspondência entre português, LIST e o sistema de notação em palavras proposta em Felipe (1998), utilizados para o protótipo:

| português | Felipe (1998) | LIST |
|----------------------------------|--------------------------|----------------------------------|
| casa | CASA | casa |
| estudei | ESTUDAR estudar | estudar |
| cortar | CORTAR-COM-FACA | cortar-com-faca |
| zebra | CAVALO^LISTRA | cavalo^listra |
| João | J-O-Ã-O | JOÃO |
| amiga | AMIG@ | amigo |
| árvores | ÁRVORE+ | árvore.@plural |
| colocar | coisa-arredondadaCOLOCAR | colocar{genobj:coisaarredondada} |
| perguntar (você para eles) | 2sPERGUNTAR3p | perguntar{psuj:2s pobj:3p} |

Tabela 1 - Correspondência entre os vocabulário do português, notação de Libras e LIST

7.3. Corpus

Para o desenvolvimento do protótipo, decidiu-se pelo trabalho em um *corpus* constituído, inicialmente, por histórias infantis¹⁷. Entre as histórias em quadrinho, optou-se pelas tirinhas semanais da Turma da Mônica, editadas pela Editora Globo, disponíveis no Portal da Mônica, em <http://www.portaldamonica.com.br>.

Selecionou-se uma só história – a de número 74 – que consistia em um panfleto instrutivo a respeito dos cuidados para se evitar a disseminação da dengue. A história é apresentada no Anexo deste capítulo. As 12 sentenças da história foram originalmente transcritas e analisar a partir do sistema de transcrição de Felipe (1998) adaptada para a transcrição LIST. A transcrição completa e análise do *corpus* foi a seguinte

Turma da Mônica - Página Semanal 74

Mônica:

- DENGUE^{tópico} vamos vencer(suj>1p,pac>3s)@imperativa

Geralmente há topicalização na LIBRAS, por se tratar de uma língua que se usa na modalidade falada, sempre em contexto, ficando a estrutura frasal O,SV ou O, V. A marca da topicalização também está na expressão facial enfática e através de uma pausa média. Poderemos marcar 3 tipos de pausa na conversação:breve, média e longa. A breve se daria entre os constituintes de um mesmo sintagma, a média entre sintagma, topicalização no final de frases e, a longa no final de períodos.

¹⁷ Além desse corpus utilizado nesse protótipo, os corpora dessa pesquisa foram analisados e foram elaboradas as regras fono-morfossintático-discursivas, repassadas e discutidas com as outras equipes pela coordenadora GT-FENEIS responsável pela articulação entre as três equipes com relação às questões linguísticas para descrição da Libras.

Existe o verbo ir e uma variante “vamos” que é usada somente em frases imperativas.
Exemplos:
hoje eu ir cinema; você precisar ir casa^estudo; ela sempre ir trabalho metrô;
vamos almoçar agora! vamos acabar trabalho!

- Água largado mosquito DENGUE surgir desenvolver. Por-isso caixa água precisar fechar-tampa(obj>coisa-achatada)
- cuidado@imperativa vaso flor, garrafa, ferro(obj>coisa-redonda), pneu velho
- tudo água dentro
- SI você febre alto, músculo dor, cabeça-dor, joelho, cotovelo, dor, MAL corpo todo, procurar médico

Cascão:

- Falar já água perigoso @enfática

Mônica:

- Só água largado @enfática água largado @enfática

Aspectos gramaticais:

1. Estrutura fasal: SVO (neutra) ou O,SV e O, V (topicalização); O_iSV_i ou O_iV_i (verbos com concordância de locativo);
2. Tipos de frase: traços não manuais – a frase enfática e a imperativa também são marcadas pela expressão facial;
3. Período composto por coordenação: frases estão justapostas, uma após a outra, geralmente sem conectivos;
4. Período composto por subordinação: aparece conectivos: por-isso, SI, depois, porque, etc;
5. Marca para tempo verbal: muitas vezes o sinal “já” é marca de passado.

Essa versão produzida pelo GT-Feneis foi revista pelo GT-NILC, que gerou as representações LIST e UNL desejáveis (ou possíveis) para cada um dos enunciados apresentados. A partir dessas representações, foram produzidos os recursos linguísticos (dicionários e gramáticas) necessários para que pudesse ser emulado computacionalmente o comportamento humano. A codificação completa do *corpus* foi a seguinte:

| Original | UNL | LIST |
|--|--|---|
| Vamos combater a dengue! | obj(fight(icl>do).@exclamation.@entry.@imperative, dengue(icl>disease).@def) agt(fight(icl>do).@exclamation.@entry.@imperative, we) | [DENGUE vamos vencer{pobj:3s psuj:1p}]{excl imp} |
| O mosquito da DENGUE nasce e se desenvolve em água parada. | mod(mosquito(icl>insect).@def, dengue(icl>disease).@def) and(develop(icl>occur), born(icl>occur).@entry) mod(water(icl>matter), stagnant(mod<thing)) | [água largado DENGUE mosquito surgir desenvolver] |

| | | |
|---|--|---|
| | plc(be born(ic>occur).@entry, water(ic>matter)) obj(be born(ic>occur).@entry, mosquito(ic>insect). @de f) | |
| Por isso, temos que manter a caixa d'água fechada. | obj(keep(ic>be).@exclamation.@entry.@need, water tank(ic>container).@def) aoj(closed(mod<thing), water tank(ic>container).@def) aoj(keep(ic>be).@exclamation.@entry.@need, we) man(keep(ic>be).@exclamation.@entry.@need, thus(ic>how)) | [por-isso caixa água precisar fechar-com-tampa{g entampaobj: coisa-achatada}]{excl} |
| Tomar cuidado com vasos... | obj(take(ic>do).@entry, care) obj(with(ic>how), flower pot(ic>container).@pl) man(take(ic>do).@entry, with(ic>how)) | [cuidado{excl} vaso flor] |
| ...garrafas, latas... | and(can(ic>container).@pl, bottle(ic>container).@entry.@pl) | [garrafa ferro{gen:coisa- redonda}] |
| ...pneus velhos! | mod(tire(pof>automobile).@exclamation.@entry.@pl, old(mod<thing)) | [pneu velho]{excl} |
| Tudo que acumule água! | od(fever(ic>disease), high(mod<thing)) mod(pain(ic>disease):01.@pl, muscle) mod(pain(ic>disease):02.@pl, head(pof>body)) plc(pain(ic>disease):03.@pl, knee(pof>body).@def.@pl) plc(pain(ic>disease):04.@pl, elbow(pof>body).@def.@pl) mod(bad feeling(ic>disease), general(mod<thing)) and(bad feeling(ic>disease), pain(ic>disease):04.@pl) and(pain(ic>disease):04.@pl, pain(ic>disease):03.@pl) and(pain(ic>disease):03.@pl, pain(ic>disease):02.@pl) and(pain(ic>disease):02.@pl, pain(ic>disease):01.@pl) and(pain(ic>disease):01.@pl, fever(ic>disease)) obj(feel(ic>be).@entry, fever(ic>disease)) aoj(feel(ic>be).@entry, you) man(feel(ic>be).@entry, if(ic>how)) | [SI você [[fever(ic>disease) alto] [músculo dor] [dor-de-cabeça] [joelho dor] [cotovelo dor] [MAL corpo todo]]{enum}] |
| procure serviços médicos | mod(service.@pl, medical(mod<thing)) obj(search(ic>do).@entry, service.@pl) | [procurar médico] |

| | | |
|--------------------------------------|--|---|
| eu falei que água é um perigo! | aoj(danger.@indef, water(ic>matter)) obj(say(ic>do).@exclamation.@entry, danger.@indef) agt(say(ic>do).@exclamation.@entry, I) | [eu falar água perigoso]{excl} |
| água parada, Cascão! | mod(water(ic>matter).@exclamation.@entry , stagnant(mod<thing)) and(Cascão(ic>person).@vocative, water(ic>matter).@exclamation.@entry) | [água largado [Cascão]{voc}]{excl} |
| água parada! | mod(water(ic>matter).@exclamation.@entry , stagnant(mod<thing)) | [água largado]{excl} |

7.4. LIST – Resultados parciais

A partir da experiência em transcrição LIST, além de um conjunto pequeno de *headwords* já estabelecido, foram estipuladas as seguintes determinações, relativas a atributos:

- **quanto à força ilocucionária:** os traços (atributos booleanos) “excl”, “imp”, “int” e “neg” são apresentados por sentenças/sinais com padrão entoacional exclamativo, imperativo, interrogativo e negativo, respectivamente. Uma sentença é afirmativa se e somente se não apresenta nenhum desses traços;
- **quanto à concordância:** de acordo com as relações de concordância verificadas para um verbo da Libras, este deve especificar zero ou mais dos seguintes atributos: “psuj/pobj” (pessoa do sujeito/objeto), “gensuj/genobj/gentampaobj” (gênero do sujeito/objeto/tampa do objeto). Os atributos de pessoa assumem valores combinando pessoa (“1”, “2” ou “3”) e número (“s(ingular)” ou “p(lural)”). Observaram-se no corpus os seguintes gêneros: “coisa-redonda” e “coisa-achatada”. Ex.:

[DENGUE vamos vencer{pobj:3s psuj:1p}]{excl imp}

“Vamos vencer a dengue!”

[por-isso caixa água precisar fechar-com-tampa{gentampaobj:coisa-achatada}]{excl}

“Por este motivo, nós temos que manter a caixa d’água fechada!”

- **quanto à “autoconcordância”:** observou-se um sinal que “concordava” com a própria idéia que exprimia, o que denominamos *autoconcordância*. Trata-se, evidentemente, de um sinal que consideramos como parametrizado, por identificar um morfema intercambiável na sua constituição. Os sinais autoconcordantes em gênero devem especificar o atributo “gen”. Ex.:

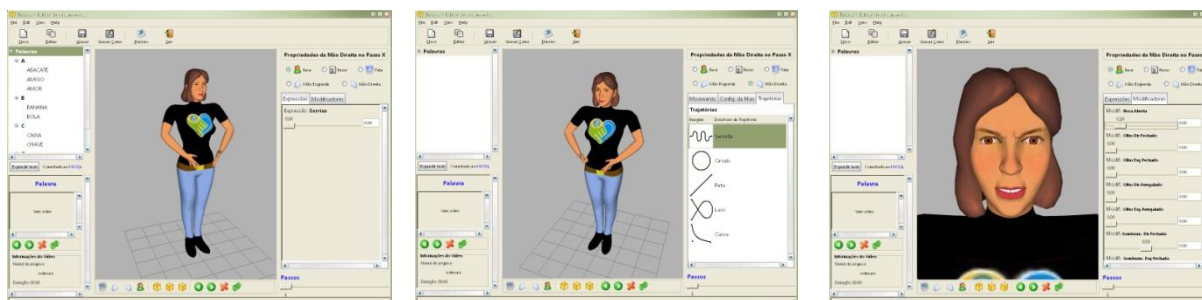
[garrafa ferro{gen:coisa-redonda}]

“... garrafas, latas ...”

- **quanto a vocativos:** embora vocativos não sejam frequentes em Libras, optamos por mantê-los em nossas versões LIST de sentenças em português. Para que, no entanto, possam ser facilmente identificados e devidamente executados ou até apagados pela equipe de síntese, o vocativo deve constituir um grupo rítmico com o traço “voc”. Ex.: “[água largado [cascão] {voc}] {excl}”
“Água parada, Cascão!”
- **quanto a enumerações:** enumerações longas costumam apresentar um padrão rítmico próprio também em libras; portanto, definimos que a enumeração longa (de três elementos ou mais) deve constituir um grupo rítmico distinto, com o traço “enum” e composto exclusivamente de tantos (sub)grupos rítmicos quantos forem seus elementos. Ex.:
[SI você [[febre alto] [músculo dor] [dor-de-cabeça] [joelho dor]
[cotovelo dor] [MAL corpo todo]] {enum}]
“Se você tiver febre alta, dores musculares, dores de cabeça, dores nos joelhos, dores nos cotovelos, e mal estar geral...”

7.5 O módulo de síntese da fala: produção do avatar

A partir do trabalho de descrição dos 8.000 sinais do Dicionário da Libras, realizado pelo GT-Feneis e das orientações da coordenadora do GT-Feneis, com relação ao projeto para o software que emularia o avatar a partir do nível fonológico da Libras, o GT-AcessibilidadeBrasil criou um protótipo para a sinalização computadorizada de cada sinal por um avatar:¹⁸



CONSIDERAÇÕES FINAL

A partir dessa exposição, fica evidente a complexidade da tarefa executada para a tradução de português para uma transcrição da Libras suficiente para alimentar um

¹⁸ Ver projeto e referências em: Tradutor Libras: Projeto TLIBRAS - Tradutor Português x LIBRAS (Língua Brasileira de Sinais) - <http://www.acessobrasil.org.br/index.php?itemid=39>

sintetizador de fala, embora nesse trabalho não tenha sido detalhadas todas as atividades específicas para a criação do PULØ e do avatar, uma vez que o foco foi apenas apresentar os sistemas de transcrição e o sistema de codificação utilizados nessa pesquisa.

A estratégia de prototipação, focando uma única história, mostrou-se apropriada, embora pareça ser bastante restritiva, uma vez que em um corpus maior de sentenças em português deverão surgir outros problemas a serem resolvidos.

Os próximos passos deverão, assim, considerar a extensão do corpus tratado, a superação das limitações já identificadas e, principalmente, a integração do sistema PULØ com o módulo de síntese de fala e a extensão do vocabulário LIST que já pode contar com 8.000 sinais que foram catalogados através do Dicionário da Libras, um subproduto desse projeto¹⁹

A análise dos sistemas para representação de uma língua de sinais em projetos para a criação de tradutores apontou para a preferência na utilização de sistemas de notação em glosa para a transcrição de dados a serem tratados em sistemas de codificação. Esse fato pode estar relacionado a questões de natureza computacional porque esses sistemas têm utilizado glosas em língua inglesa, além de sinais gráficos e tipográficos, o que induz à utilização de transcrições lineares em glosa e não de escritas ou transcrições icônicas e/ou pictóricas que apresentam dificuldades para serem utilizadas pelos sistemas de decodificação existentes.

Ao se analisar a produção de textos sinalizados por avatar, em sua maioria, o processo ainda está em uma fase de tradução semiautomática, auxiliada por humano, porque os sinais têm sido gerados sem uma interface entre o sistema de decodificação do tradutor com o seu sistema de sintetização da fala (programação gráfica para a sinalização baseada no nível fonológico) que precisam se conectar automaticamente para a geração de textos sinalizados por avatar.

Portanto, o problema a ser enfrentado é: como se conectarão essas duas partes do processo de tradução para se gerar frases em língua de sinais através de avatar inteligente que decodificará o nível fonológico das frases fornecidas pelo sistema de codificação do tradutor?

REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, W. M. Sistema de transcrição da língua brasileira de sinais voltado a produção de conteúdo sinalizado por avatares 3D. Tese de Doutorado. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação -Campinas, SP:[s.n.], 2012. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?code=000881689> Acesso em dezembro/2012
- BATTISON, R. Lexical borrowing in American Sign Language. Silver Spring, MD: Linstok, 1978.
- BELLUGI, U. & KLIMA, E.S. The acquisition of three morphological systems in American Sign Language. Papers and Reports on Child Language Development, 21, 1982: 1-34

¹⁹ Dicionário distribuído pelo INES e disponível no site: www.librasencontextro.org

BÉBIAN, R.A.A. *Mimographie, ou essai d'écriture mimique propre à régulariser le langage des sourds-muets*. Paris: Louis Colas. 1825

----- *Essai sur les sourds-muets et sur le langage naturel ou introduction à une classification naturelle des idées avec leurs signes propres*. Paris : J.G. Dentu. 1817

BERENEZ, N. *Person and deixis in Brazilian Sign Language*. Dissertation for Doctor of Philosophy in Linguistics, University of California, Berkeley. 1996

BERENEZ, N. & FERREIRA BRITO, L. Pronouns in BCSL and ASL. *Sign Language Studies*, 87. 1990: 26-36.

CALDAS, B.F. *Narrativas em LSCB: Um estudo sobre referência*. Dissertação de Mestrado. Universidade Federal do Rio de Janeiro. 1992

CHRISTMANN, K.E et al. O software ELAN como ferramenta para transcrição, organização de dados e pesquisa em aquisição da língua de sinais. *Anais do IX Encontro do CELSUL*. Universidade do Sul de Santa Catarina. Palhoça, 2010. Disponível em:
<http://celsul.org.br/Encontros/09/artigos/Karina%20Christmann.pdf> Acesso em 10 de fevereiro de 2014.

CORRADI, J. A. M. *Ambientes informacionais digitais e usuários surdos: questões de acessibilidade*. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. 2007. Disponível em:
http://www.marilia.unesp.br/Home/Pos Graduacao/CienciadaInformacao/Dissertacoes/corradi_jam_me_mar.pdf

DENARDI, R. M. et al. AGA-Sign: animador de gestos aplicado à língua de sinais. *INFOCOMP: Journal of computer science*, v. 4, n. 1, p. 46-52, 2005. Disponível em:
<http://www.dcc.ufla.br/infocomp/artigos/v4.1/art06.pdf> Acesso em: 20 out. 2014.

FARIA, C. V. S. (1995) *Atos de fala: O Pedido em Língua Brasileira de Sinais*. Dissertação de Mestrado. Rio de Janeiro: UFRJ.

FELIPE, T. A. O discurso verbo-visual na língua brasileira de sinais – Libras. *Bakhtiniana. Revista de Estudos do Discurso*. V. 8, n. 2, 2013: 67-99. Disponível em:
<http://revistas.pucsp.br/index.php/bakhtiniana/article/view/14141>

----- *Bilinguismo e Educação Bilíngue: questões teóricas e práticas pedagógicas* Fórum Permanente do INES. Rio de Janeiro: INES, Divisão de Estudos e Pesquisas, 2012: 7-23

----- *O processo de formação de palavras na LIBRAS*. ETD – Educação Temática Digital, Campinas, V.7, N.2, 2006:199-216. Disponível em:
<http://www.fae.unicamp.br/revista/index.php/etd/article/view/1642/1489>

----- GP-Libras-FENEIS. *Dicionário da Libras*. CD-ROM. 2005. Disponível em:
<http://www.librasemcontexto.org/>

----- *LIBRAS em contexto - Curso Básico - Livro do estudante/Cursista*. CDU. Brasília: MEC - SEESP - Programa Nacional Interiorizando a Libras, 7ª Edição. Disponível em:
<http://www.librasemcontexto.org/>

----- *Introdução à gramática da LIBRAS*. In: MEC/SEESP. (Org.). *Educação Especial - Língua Brasileira - Série Atualidades Pedagógicas - 4. 2ª ed*. Brasília, V. III, 1999, p.81-123. Disponível em:
http://www.librasemcontexto.org/producao/lingua_brasileira_sinais_mec.pdf

----- *A relação sintático-semântica dos verbos na Língua Brasileira de Sinais (LIBRAS)*. Tese de Doutorado em Linguística e Filologia. Universidade Federal do Rio de Janeiro. Rio de Janeiro. V.1: 01-143; V.2: 144-298; V.3: Transcrição dos dados das Fitas: 1,2,3,4,5,6 (Banco de dados). 1998.

----- *A Valência dos Verbos na LSCB" nos Anais do II Congresso da Associação de Estudos da Linguagem do Rio de Janeiro (ASSEL)*. Faculdade de Letras da UFRJ. 1993: 216-231

----- *Por uma tipologia dos verbos da LSCB*. *Anais do VII Encontro Nacional da ANPOLL*. Goiânia: Vol 2 - Linguística, 1993: 724-744

----- *A relação sintático-semântica dos Verbos da LSCB" nos Anais do VII Encontro Nacional da ANPOLL*, no GT Linguagem e Surdez, realizado de 17 a 20 de maio de 1992 em Porto Alegre

----- *Aquisição de linguagem por crianças surdas*. Monografia para conclusão da disciplina Psicolinguística no curso de Doutorado em Linguística, UFRJ, Rio de Janeiro. 1991

----- *Coesão Textual em Narrativas Pessoais na LSCB*. Monografia de conclusão da disciplina "História da Análise do Discurso". Curso de Doutorado em Linguística. Rio de Janeiro: UFRJ. 1991

----- *Do Discurso à gramática da LSCB*. Seminário sobre FUNCIONALISMO EM CURSO. Rio de Janeiro: UFRJ, 1991: 52-55.

----- *A Estrutura Frasal na LSCB*. *Anais do IV Encontro Nacional da ANPOLL*. Recife, 1989: 663-672.

----- *O Signo Gestual-Visual e sua Estrutura Frasal na Língua dos Sinais dos Centros Urbanos Brasileiros*. Dissertação de Mestrado, UFPE, PE. 1988

FERREIRA BRITO, L. *Classificadores em LSCB*. *Anais do IV Encontro Nacional da ANPOLL*. Recife, 1989: 640-654.

- FERREIRA BRITO et al. (1995) Por uma gramática de Língua de Sinais. Rio de Janeiro: Tempo Brasileiro: UFRJ, Depto. Linguística e Filologia.
- FERREIRA BRITO, L & LANGEVIN, R. Negação em uma Língua de Sinais Brasileira. Delta, Vol. 10, no 2: 1994: 309-327.
- FISCHER, R. Language of action. En Fischer, R. & Lane, H. (eds.) Looking back: A Reader of the History of Deaf Communities and Their Sign Languages 1993: 429-455). Hamburg: Signum.
- The notation of sign languages: Bébien's mimographie. Bos, H. F., Schermer, G. & Gertrude, M. (eds.). Sign Language Research 1994: Proceedings of the Fourth European Congress on Sign Language Research Hamburg: Signum. 1995: 285-302.
- GAMA, F. J. (1875) Iconographia dos Signaes dos Surdos-Mudos. Rio de Janeiro: Typographia Universal de E. & H. Laemmert.
- HOEMANN, H.W., OATES, E., HOEMANN, S. Linguagem de sinais do Brasil. Centro Educacional para Deficientes Auditivos. Porto Alegre. RS. 1983.
- HULST, Harry van der. Units in the analysis of signs. In: Phonology 10: 2, 1993:209-241.
- HULST, H.G. van der and CHANNON, Rachel. Notation systems. In: Diane Brentari (ed.). Sign Language, Cambridge Survey of Sign Linguistics and Sign Languages. 2010: p. 151-17
- KARNOPP, L. B. Literatura Surda. Coleção Letras Libras. UFSC. 2008. Disponível em: http://www.libras.ufsc.br/colecaoLetrasLibras/eixoFormacaoEspecifico/literaturaVisual/assets/369/Literatura_Surda_Texto-Base.pdf
- Aquisição do parâmetro configuração de mão na língua brasileira dos sinais (LIBRAS): estudo sobre quatro crianças surdas, filhas de pais surdos. Dissertação de Mestrado. Porto Alegre: PUC. 1994.
- MARTINS, R., PELIZZONI, J., HASEGAWA, R. PULØ - Para um sistema de tradução semi-automática português-libras. XXV Congresso da Sociedade Brasileira de Computação. UNISINOS. São Leopoldo/RS. 2005 Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/til/2005/0014.pdf>. Acesso em dezembro de 2005
- MCCLEARY, L. & VIOTTI, E. Transcrição de dados de uma língua sinalizada: um estudo piloto da transcrição de narrativas na língua de sinais brasileira (LSB). In H. Salles (Org) Bilinguismo dos surdos: questões linguísticas e educacionais. Goiania: Canone Editorial. 2007:73-96.
- MCCLEARY, L; Viotti, E e LEITE, T.A. Descrição das línguas sinalizadas: a questão da transcrição dos dados. Alfa, São Paulo, 54 (1): 265-289, 2010. seer.fclar.unesp.br/alfa/article/download/2880/2654 Acesso em 10 de fevereiro de 2014.
- NASCIMENTO, V. Contribuições bakhtinianas para o estudo da interpretação da língua de sinais. Tradterm, v. 21, p. 213-236, 2013. Disponível em: http://myrtus.uspnet.usp.br/tradterm/site/images/revistas/v21n1/13_vinicius21f.pdf Acesso em set/2013.
- NEIDLE, C. SignStream Annotation: conventions used for the American Sign Language Linguistic Research Project. 2002. Disponível em: <http://www.bu.edu/asllrp/>. Acesso em: 17 março 2010
- OATES, E. Linguagem das Mãos. Rio de Janeiro: Gráfica Ed. Livro. 1969
- OVIEDO, A. Las "huellas" de la Mimographie (Bébien 1825) en el sistema de transcripción de las señas de William C. Stokoe. Lenguaje, 2009, 37 (2), 293-313
- Roch Ambroise Auguste Bébien (*1789 / †1839). Pionero de los estudios sobre la sordera. 2007. Disponível em: <http://www.cultura-sorda.eu/4.html> Acesso em fevereiro de 2012.
- PADDEN, C. (1983) Iteration of Morphology and Syntax in American Sign Language, Thesis document, University of California, San Diego.
- QUADROS, R.M. As categorias Vazias Pronominais: uma Análise Alternativa com Base na LIBRAS e Reflexos no Processo de Aquisição. Dissertação de Mestrado. Porto Alegre: PUC. RS. 1995
- RODRIGUES, Carlos Henrique. Efeitos de modalidade no processo de interpretação simultânea para a Língua de Sinais Brasileira. ReVEL, v. 10, n. 19, 2012. [www.revel.inf.br]. www2.fcsh.unl.pt/giud-clun1/civ/2.4-ELAN.pdf Acesso em 10 de fevereiro de 2014.
- STOKOE, W.C. Sign Language Structure. An Outline of the Visual Communication Systems of the American Deaf. Silver Spring: Linstok Press. (1993 [1960]).
- STOKOE, W.C. (1960) Sign Language Structure: An outline of the visual communication systems of American Deaf in Studies in Linguistics: Occasional paper, 8.
- STOKOE, W.C., Casterline, D. & Croneberg, C. A Dictionary of American Sign Language on Linguistic Principles. Silver Spring: Linstok Press. 1965.
- SUTTON, V. Researcher's Resources SignWriting. Sign Language & Linguistics Vol. 2, no. 2, John Benjamins, 1998: 271-281.