

A Process for Semi-Automated Construction of Sign Language Dictionaries

Marília Soares

Universidade Federal da Paraíba
marilia.soares@lavid.ufpb.br

Beatriz Meireles

Universidade Federal da Paraíba
beatriz.meireles@lavid.ufpb.br

Sandro Gonçalves

Universidade Federal da Paraíba
sandro.goncalves@lavid.ufpb.br

Manuella Aschoff

Universidade Federal da Paraíba
manuella.lima@lavid.ufpb.br

Tiago Maritan

Universidade Federal da Paraíba
maritan@lavid.ufpb.br

Valdecir Becker

Universidade Federal da Paraíba
valdecir@lavid.ufpb.br

ABSTRACT

In the scientific literature, there are several solutions developed for machine translation of digital contents into sign languages. Generally, these solutions aim to reduce communication and access to information barriers for deaf in Information and Communication Technologies (ICTs). One of the main challenges of this type of system is the creation of visual representations (animations or videos) for the signs of the language, generally thousands of signs, which is a very time-consuming task. Thus, in this work, we propose a process (or methodology) for the semi-automatic construction of sign language dictionaries. This process combine automatic tasks, such as motion capture (e.g., using Microsoft Kinect) with manual tasks involving 3D animators and sign language specialists. To evaluate this proposal, a case study was developed for the Suíte VLibras platform, an open-source machine translation platform to Libras. The result was the creation of a Libras dictionary with more than 13,000 signs which is currently used in more than 1500 websites.

KEYWORDS

Sign Language Dictionary; Signs Dictionary Management System

1 INTRODUÇÃO

Pessoas surdas comunicam-se naturalmente usando línguas gestuais, que são percebidas pelos olhos e se realizam no espaço com articuladores visuais (as mãos, o corpo, os movimentos e o espaço de sinalização), estas línguas são denominadas línguas de sinais (LS). Apesar das LS, em muitos países, serem reconhecidas como língua, os problemas enfrentados pelos surdos no tocante ao acesso à informação permanecem. Em consequência disso, observa-se uma grande dificuldade dos surdos exercerem a sua cidadania e terem acesso à informação através das LS, o que acaba implicando em atraso linguístico, dificuldades sociais, de relacionamento e de aquisição do conhecimento.

Visando minimizar a marginalização dos surdos, vários trabalhos relacionados à tradução automática de conteúdos em línguas orais para línguas de sinais vêm sendo desenvolvidos. Dentre estes trabalhos, elencam-se os de Araújo et al [2, 3], Huenerfauth [7–9], López-Ludeña et al [10, 11], Morrissey [12], dentre outros.

In: XIV Workshop de Trabalhos de Iniciação Científica (WTIC 2017), Gramado, Brasil. Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres. Porto Alegre: Sociedade Brasileira de Computação, 2017.

© 2017 SBC – Sociedade Brasileira de Computação.
ISBN 978-85-7669-380-2.

Nessas soluções, a sinalização do conteúdo em LS geralmente usa avatares 3D, que utilizam animações ou vídeos para renderizar as mensagens.

Contudo, a construção das animações ou vídeos dos sinais é geralmente um processo muito caro e trabalhoso, pois criar um dicionário de sinais correto e amplo, com dezenas de milhares de sinais, pode envolver muitos recursos tecnológico, logísticos e também o esforço coordenado de uma equipe multidisciplinar. Portanto, trata-se de um dos maiores desafios para o desenvolvimento desse tipo de plataforma de tradução automática.

Para reduzir essas dificuldades, o presente trabalho propõe a definição de um processo para construção de dicionário em LS. A proposta combina etapas automáticas, como a captura digital dos sinais usando sensores de movimento, com etapas manuais de ajuste, adaptação, revisão e homologação desses sinais com animadores 3D e especialistas em LS. Para validar a proposta, um estudo de caso foi conduzido na construção do dicionário de um importante tradutor automático para Língua Brasileira de Sinais (LIBRAS). O resultado foi a criação de um dicionário com cerca de 13.000 sinais e que está disponível em mais de 1500 websites.

O artigo está organizado da seguinte forma. Na seção 2 são descritos alguns trabalhos relacionados à criação de dicionários em LS, enquanto que na seção 3 é apresentada a solução proposta contendo a metodologia de construção de dicionários para LS. Em seguida, na seção 4, encontra-se a prova de conceito de aplicação desse processo na construção do dicionário da Suíte VLibras. Finalmente são demonstrados alguns resultados e os trabalhos futuros.

2 TRABALHOS RELACIONADOS

Bragg et al. [1] apresenta um dicionário online, denominado ASL-Search, que é alimentado por seus próprios usuários. O ASL-Search utiliza Análise Semântica Latente (Latent Semantic Analysis - LSA), que tradicionalmente modela relações entre palavras e documentos para modelar relações entre características e sinais em um banco de dados de consultas de usuários baseadas em recursos. Esta proposta é feita por Bragg et al [1] como forma de atender variabilidade dos sinais.

Wilcox et al. [13] apresentam uma proposta de dicionário que é desenvolvido para funcionar como bilíngue (ASL-Inglês). Ele contém sinais de ASL em vídeo digital, cujos movimentos usam a tecnologia QuickTime da Apple. Este dicionário permite a busca por sinais ASL através da pesquisa direta, usando aspectos formais da ASL e, por pesquisas mistas, usando o inglês e a ASL. Além disso, ele serve como motor central para o MM-DASL, que é um

WebMedia'2017: Workshops e Pôsteres, WTIC, Gramado, Brasil

aplicativo que pode ser utilizado em qualquer LS, pois permite que pesquisadores e desenvolvedores o usem para criar seus próprios dicionários em LS.

Buttussi et al. [5] propõe um dicionário on-line internacional de linguagem gestual, chamado 3DictSL. Diante disso propõe uma arquitetura cliente-servidor, que inclui ferramentas de criação, com o suporte a novos idiomas de sinais, para estender o dicionário tornando-o mais robusto e adaptado para diferentes LS. No entanto, o dicionário pode ser visualizado apenas por navegadores que suportem as tecnologias X3D, H-Anim e Javascript SAL.

Duduchi e Capovilla [4] apresentam uma solução computacional chamada BuscaSigno, que utiliza um sistema de busca com objetivo de permitir que o usuário localize diretamente qualquer sinal da LIBRAS por meio da seleção, por mouse ou teclado, e a partir de um menu de busca quiriêmica¹.

Contudo, embora esses trabalhos, permitam a criação e busca de sinais em dicionários, nenhum deles propõe um processo ou metodologia para construção desses dicionários. Com isso, não há um método para escolha/pesquisa dos sinais que selecionados, e como consequência, se esses dicionários envolverão um conjunto representativo dos sinais da língua. Além disso, eles também não definem uma metodologia para animação, avaliação e gerenciamento desses dicionário de sinais. O preenchimento dessas lacunas é a proposta do processo de construção e gerenciamento de dicionários em línguas de sinais que está sendo proposto e será apresentado na Seção 3.

3 PROCESSO DE CONSTRUÇÃO DE DICIONÁRIOS EM LÍNGUA DE SINAIS

A construção de um dicionário seja ele de língua oral (LO) ou LS, não é algo trivial. Em se tratando de um dicionário de LS, a atividade é ainda mais complexa, pois é necessário muita pesquisa, dedicação e uma equipe multidisciplinar, visto que o canal de comunicação destas línguas são gesto-visual exigindo, portanto, além de linguístas bilíngue da área, pessoal capacitado para lidar com tecnologias multimídias.

Neste sentido o presente artigo visa contribuir com a proposta de um processo para construção de dicionários em LS. A metodologia proposta encontra-se esquematizada na Figura 1.

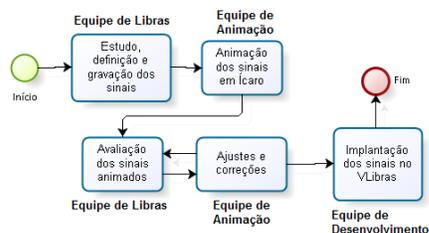


Figura 1: Fluxograma de Criação de um Dicionário LO/LS

De acordo com a Figura 1, de maneira geral, o processo funciona da seguinte forma. Inicialmente uma etapa de "Pesquisa de Sinais" é

¹Partes mínimas de um sinal. Os quiriêmas se equiparam aos fonemas das línguas orais

realizada com o objetivo de definir um conjunto de sinais representativos e que esteja de acordo com a norma culta da LS. Em seguida, os sinais são gravados e capturados ("Captura de Sinais") com o apoio de sensores de movimento, como por exemplo, o Microsoft Kinect. Essa etapa é importante porque ela reduz o esforço de animação utilizado na próxima fase do processo, além de garantir uma maior naturalidade dos movimentos sinalizados. Em seguida, os sinais são animados e ajustados por uma equipe de animadores 3D, com base na captura realizada na etapa anterior e nos vídeos de referência gravados. Em seguida, os sinais passam por um ciclo de Avaliação e Correção, até que sejam aprovados por uma equipe de especialistas em LIBRAS. Por fim, eles são convertidos para um formato compatível a com a plataforma de tradução automática e implantados na mesma. Este processo é melhor detalhado na figura 1

3.1 Pesquisa e Captura de Sinais

Os sinais das LS até pouco tempo atrás apenas eram ensinados entre gerações, não possuíam uma forma de registro e catalogação. Com o advento das Tecnologias da informação e comunicação (TICs) esses registros foram possíveis, posto que a captura e armazenamento de multimídias ficou mais acessível à população em geral.

Sendo assim, o primeiro passo da pesquisa de sinais é o resgate destes junto as comunidades surdas seguido do registro em vídeos. Atualmente muitos desses registros são disponibilizados na internet através de serviços de vídeo sob demanda (Ex: Youtube). No entanto, quando necessário, essa comunidade cria sinais para palavras da LO cuja representação em LS ainda é inexistente. No que tange a construção de dicionários de sinais é imprescindível criar-se a correlação entre as palavras da LO e os sinais da LS para que sejam garantidas a semântica e pragmática da tradução.

Os processos de pesquisa e captura de sinais são uma tarefa realizada inerentemente por especialistas na LS, sejam eles surdos ou intérpretes. Essas atividades envolvem uma catalogação dos sinais existentes na LS alvo e esta deve ser registrada em vídeos e/ou animações computacionais. Sendo assim, os especialistas na LS fazem o levantamento dos sinais, realizam o registro visual dos mesmos e capturam os movimentos utilizando um sensor de movimento.

A pesquisa de sinais inicia-se pelo levantamento dos sinais mínimos da língua, ou seja, as letras que compõem o alfabeto, e só então busca-se os sinais que representam palavras completas. Este levantamento é essencial diante da possibilidade de se realizar a datilologia² de qualquer sinal inexistente no dicionário, o que pode afetar diretamente a qualidade das traduções realizadas.

Após o levantamento dos sinais eles são registrados em vídeos por especialistas surdos. Nestes vídeos são inseridas as legendas, em LO, com as devidas identificações e estes passam a ser usados como vídeos de referência no processo de animação dos sinais detalhado na sessão 3.2.

Concomitantemente ao processo de gravação dos vídeos de referência é realizada a captura e rastreamento dos movimentos utilizando um sensor de movimento que tem como objetivo automatizar o processo de animação dos sinais. Atualmente este processo é realizado de forma semi-automática visto que os principais sensores de

²Soletização manual

A Process for Semi-Automated Construction of Sign Language Dictionaries

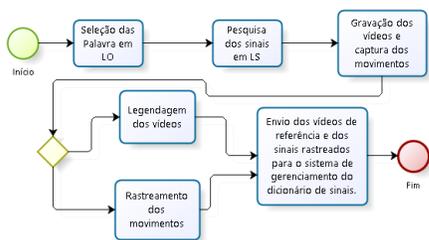


Figura 2: Fluxograma para a Pesquisa e Captura dos Sinais

movimento disponíveis no mercado, ainda não conseguem perceber detalhadamente as expressões faciais e corporais, essências para as LS.

3.2 Criação dos sinais

Nessa fase, uma equipe de animadores 3D recebe os movimentos dos sinais rastreados usando sensores de movimento, com os vídeos de referência dos sinais produzidos na etapa anterior e faz as adaptações no modelo do agente animado virtual utilizado pela plataforma de tradução automática. De forma geral, este tipo de animação é feita usando avatares 3D que possuem todas as suas articulações modeladas para permitir a reprodução do movimento de sinais com a maior naturalidade possível. Nesse contexto, a utilização de sensores de movimento auxilia também no processo de animação, uma vez que os movimentos capturados e rastreados são geralmente mais naturais do que aqueles criados manualmente por animadores 3D.

Para realizar as animações dos sinais é usado um software de animação 3D. Estes softwares possuem versões gratuitas, tais como o Blender, e versões pagas, a exemplo do Maya. Os softwares de animação são ferramentas que permitem a criação de projetos em 3D, oferecendo funcionalidades para modelagem, renderização, animação, pós-produção, criação e visualização de conteúdo 3D interativo.

3.3 Avaliação e Implantação dos sinais

A avaliação dos sinais gerados é realizada pelos especialistas na LS que fizeram a pesquisa e geração dos vídeos de referência. Esta avaliação inclui a visualização da animação em busca de identificar possíveis falhas na execução do sinal, configuração de mãos, orientação da mão, expressões faciais e corporais e dos pontos de articulação das mãos com o corpo.

Após esta minuciosa verificação os sinais aprovados são incorporados ao dicionário enquanto que os sinais reprovados passarão por ciclos de correções dos erros registrados durante as avaliações. Assim, após corrigidos, são novamente avaliados até atingir o nível de aprovação.

A implantação dos sinais, por sua vez, é uma etapa do processo de construção do dicionário em que os sinais aprovados são convertidos para o formato utilizado na plataforma de tradução automática (vídeos, animações, metadados, entre outros) e adicionados a mesma.

É importante destacar que o processo proposto, além de ser utilizado para construção de um dicionário a partir do zero, também pode ser aplicado para produzir atualizações e inclusões de novos

WebMedia'2017: Workshops e Pôsteres, WTIC, Gramado, Brasil

sinais nesses dicionários. Isso é importante uma vez que as línguas de sinais são línguas naturais e vivas, e novos sinais podem surgir frequentemente.

4 ESTUDO DE CASO - SUÍTE VLIBRAS

Para validar o processo proposto, um estudo de caso foi conduzido para construção do dicionário de sinais da plataforma Suíte VLibras³. Nesse processo, a fase de pesquisa e captura dos sinais foi realizada por uma equipe de surdos e intérpretes de Libras da UFPB, do INES⁴ e da Câmara dos Deputados (parceiros do projeto). As pesquisas foram conduzidas utilizando dicionários de sinais que eram referência nos cursos de Letras, modalidade Libras (Ex: "Acesso Brasil" e Dicionário trilingue de Língua de Sinais [6]), juntamente com reuniões com grupos de surdos especialistas.

A gravação, captura e rastreamento dos sinais foi realizada utilizando o Microsoft Kinect v2 (ver Figura 4), um sensor de movimento de baixo custo. Em seguida, estes parâmetros de movimento são transportados para o modelo do agente animado virtual (avatar-3D) do VLibras que foi modelado no software Blender com uma armadura composta por 82 ossos, distribuídos da seguinte forma:

- 15 ossos em cada uma das mãos para configurar a posição dos dedos;
- 23 ossos para configurar os elementos faciais;
- 22 ossos para configurar os movimentos de braço e de corpo;
- 7 ossos auxiliares que não deformam a malha do avatar-3D diretamente.

Dessa forma, para configurar, por exemplo, os movimentos dos dedos, é necessário definir os parâmetros de localização e rotação de cada um dos 15 ossos da mão. O mesmo deve ser feito para configurar os demais ossos. Além disso, para combinar a deformação entre ossos relacionados foi utilizada cinemática inversa (inverse kinematics - ik). O modelo do avatar-3D é ilustrado na Figura 3. As Figuras 3-a, 3-b e 3-c ilustram esse modelo com ênfase nos ossos da face, das mãos e do corpo, respectivamente. A figura 3-d mostra como o avatar é após a renderização.

Após a captura e rastreamento, os sinais eram encaminhados para uma equipe de animadores 3D que utilizavam os movimentos rastreados e os vídeos de referência para animar e ajustar os sinais. Essa etapa era realizada usando o software Blender (ver Figura 5). Em seguida, os sinais animados passavam por ciclos de avaliações e correções envolvendo a equipe de surdos e intérpretes de LIBRAS das três instituições que participavam do projeto.

É importante destacar que o processo de distribuição dos sinais entre os animadores, avaliações e correções eram mediados por um sistema de gerenciamento do dicionário de sinais, desenvolvido pela equipe do projeto, denominado de DLibras. Esse sistema era um sistema Web, que possuía como recursos a possibilidade atribuir sinais para animadores, realizar ciclos de avaliação e correção dos

³ A Suíte VLibras (www.vlibras.gov.br) é o resultado de uma parceria entre o Ministério do Planejamento, Desenvolvimento e Gestão (MP), por meio da Secretaria de Tecnologia da Informação (STI) e a Universidade Federal da Paraíba (UFPB), que consiste em um conjunto de ferramentas computacionais de código aberto, responsável por traduzir conteúdos digitais (texto, áudio e vídeo) para a Língua Brasileira de Sinais - LIBRAS, tornando computadores, dispositivos móveis e plataformas Web acessíveis para pessoas surdas.

⁴ Instituto Nacional de Educação de Surdos

WebMedia'2017: Workshops e Pôsteres, WTIC, Gramado, Brasil

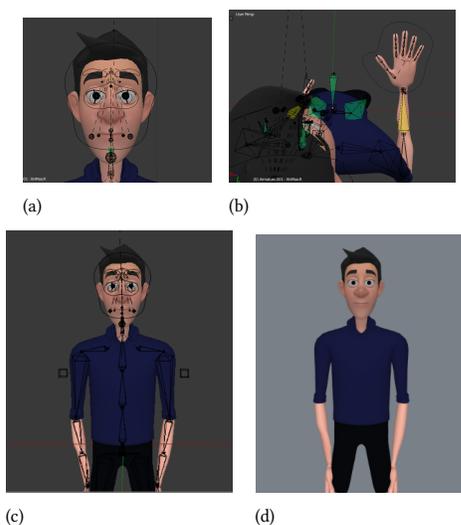


Figura 3: Modelo do Avatar-3D Ícaro



Figura 4: Gravação dos vídeos de referência e captura dos sinais

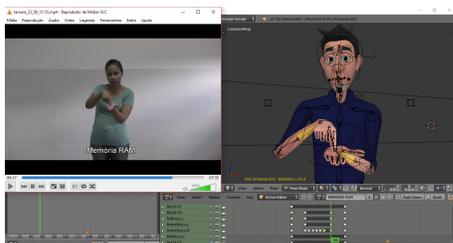


Figura 5: Processo de animação dos sinais no Blender

sinais criados, além da exportação e implantação no VLibras. Na fase de implantação, os sinais aprovados eram exportados do Blender para a plataforma Unity, utilizada pelo VLibras para renderizar e apresentar os sinais e sentenças em Libras.

Como resultado de aplicação desse processo, foi construído um dicionário com cerca de 13.000 sinais em LIBRAS. Esse dicionário foi incorporado ao sistema VLibras, e atualmente encontra-se disponível em mais de 1500 websites, dentre eles os sites do Governo Federal, Câmara dos Deputados, Senado Federal, Conselho Nacional de Justiça, entre outros. Além disso, a ferramenta conta atualmente com mais de 100 mil downloads e está disponível para os sistemas operacionais Microsoft Windows, Linux, Android e iOS, além dos navegadores Google Chrome, Mozilla Firefox e Safari.

5 CONCLUSÕES

Neste trabalho é apresentada uma proposta de processo para construção de dicionários em línguas de sinais usando avatares 3D. O desenvolvimento desses dicionários de sinais consiste num dos principais desafios relacionados ao desenvolvimento de ferramentas de tradução automática para línguas de sinais, e seu tamanho e abrangência tem bastante relação com a qualidade destes tradutores.

O processo proposto combina etapas automáticas envolvendo a captura de sinais utilizando sensores de movimento, como, por exemplo, o Microsoft Kinect v2, com etapas manuais de animação, ajustes e avaliação de sinais por uma equipe multidisciplinar. Para validar a solução proposta, um estudo de caso foi conduzido no escopo do projeto VLibras que resultou no desenvolvimento de um dicionário de cerca de 13.000 sinais, e na incorporação desses sinais na plataforma VLibras, que encontra-se disponível atualmente em mais de 1500 websites.

Como proposta de trabalho futuro, pretende-se ampliar o dicionário criado principalmente através da inserção de termos técnicos e específicos, de diferentes áreas de atuação. Além disso, almeja-se aplicar esse processo na criação de dicionários envolvendo diferentes línguas de sinais, auxiliando no processo de criação de plataformas abertas de tradução automática para múltiplas línguas de sinais.

REFERÊNCIAS

- [1] K. Rector D. Bragg and R. E. Ladner. 2015. A user-powered American Sign Language dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1837–1848.
- [2] T. M. U. de Araújo, F. L. S. Ferreira, D. A. N. dos S. Silva, F. H. Lemos, G. Pessoa Neto, D. Omaia, G. L. de Souza Filho, and T. A. Tavares. 2013. Automatic generation of Brazilian sign language windows for digital TV systems. *Journal of the Brazilian Computer Society* 19, 2 (Junho 2013), 107–125.
- [3] T. M. U. de Araújo, F. L. S. Ferreira, D. A. N. dos S. Silva, L. D. Oliveira, E. L. Falcão, L. A. Domingues, V. F. Martins, I. A. C. Portela, Y. S. Nóbrega, G. L. de Souza Filho Lima, T. A. Tavares, and A. N. Duarte. 2014. An approach to generate and embed sign language video tracks into multimedia contents. *Information Sciences* 281 (Outubro 2014), 762–780.
- [4] M. Duduchi and F. C. Capovilla. 2006. BuscaSigno: a construção de uma interface computacional para o acesso ao léxico da língua de sinais Brasileira. In *Proceedings of VII Brazilian symposium on Human factors in computing system*. 21–30.
- [5] L. Chittaro F. Buttussi and M. Coppo. 2007. Using web3d technologies for visualization and search of signs in an international sign language dictionary. In *Proceedings of the International Conference on 3D Web Technology*. 61–70.
- [6] W. D. Raphael F. C. Capovilla and a. C. L. Mauricio. 2012. *Novo Deit-Libras: Dicionário enciclopédico ilustrado trilingue da Língua de Sinais Brasileira (Libras) baseado em Linguística e Neurociências Cognitivas*. EdUSP., São Paulo.
- [7] M. Huenerfauth. 2004. A multi-path architecture for machine translation of English text into American Sign Language animation. In *Proceedings of the Student Research Workshop at HLTNAACL*. 25–30.
- [8] M. Huenerfauth. 2005. American sign language generation: multimodal NLG with multiple linguistic channels. In *Proceedings of the ACL Student Research Workshop*. 37–42.
- [9] M. Huenerfauth. 2005. Representing coordination and non-coordination in an american sign language animation. In *Proceedings of the 7th international ACM*

A Process for Semi-Automated Construction
of Sign Language Dictionaries

WebMedia'2017: Workshops e Pôsteres, WTIC, Gramado, Brasil

- SIGACCESS conference on Computers and accessibility*. 44–51.
- [10] V. López-Ludeña, C. González-Morcillo, J.C. López, E. Ferreiro, J. Ferreiros, and R. San-Segundo. 2014. Methodology for developing an advanced communications system for the Deaf in a new domain. *Knowledge-Based Systems* 52 (Janeiro 2014), 240–252.
- [11] V. López-Ludeña, C. González-Morcillo, J.C. López, E. Ferreiro, J. Ferreiros, and R. San-Segundo. 2014. Translating bus information into sign language for deaf people. *Engineering Applications of Artificial Intelligence* 32 (Junho 2014), 258–269.
- [12] S. Morrissey and A. Way. 2013. Manual labour: tackling machine translation for sign languages. *Machine Translation* 27, 1 (Março 2013), 25–64.
- [13] D. Wood D. Cokely S. Wilcox, J. Scheiban and W. C. Stokoe. 1994. Multimedia dictionary of American Sign Language. In *Proceedings of the first annual ACM conference on Assistive technologies*. 09–16.