

Investigating the Collaborative Process of Subtitles Creation and Sharing for Videos on the Web

Jessica Oliveira Brito
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, S/N
29075-910 Vitória - ES, Brazil
jeubrito@gmail.com

Rodrigo Laiola Guimaraes
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, S/N
29075-910 Vitória - ES, Brazil
rlaiola@inf.ufes.br

Celso A. S. Santos
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, S/N
29075-910 Vitória - ES, Brazil
saibel@inf.ufes.br

ABSTRACT

In this paper we concentrate on the study of the collaborative practices of enthusiasts that create and share subtitles for third-party videos. Based on preliminary results from interviews with some volunteers, we formalize the subtitles creation and sharing process using a business process management model and compare it with other collaborative and crowdsourcing models. We expect that our initial observations can bring a new understanding of the process and, thus, help in the design of next generation video enriching tools.

KEYWORDS

Captioning; subtitles; collaboration; crowdsourcing.

1 INTRODUÇÃO

O processo de anotação de conteúdo audiovisual para melhorar, ou enriquecer, a experiência do usuário final é um tema que continua a despertar interesse em multimídia [1]. Enquanto a análise de conteúdo audiovisual concentra-se predominantemente em ferramentas automáticas [20], profissionais podem fornecer anotações de mídia de alta qualidade [6]. Mesmo redes sociais *on-line* [4], e até a terceirização para uma rede indefinida (e geralmente grande) de pessoas (*crowdsourcing*) [16], também podem ser usadas no enriquecimento de conteúdo audiovisual.

Anotações de mídia podem estar associadas a arquivos de áudio, vídeo, imagem, marcações de tinta [2] ou texto. No caso deste último, podemos citar as legendas, que são um tipo de anotação de mídia que enriquece a apresentação do conteúdo audiovisual, fornecendo aos telespectadores informações de fala e eventos representativos. Legendas são particularmente importantes para incluir pessoas com perda auditiva ou surdez; porém, seu uso não está limitado a esses casos. Os benefícios do uso de legendas são amplamente relatados na literatura em

cenários que avaliam questões de acessibilidade [3][12], esforço cognitivo [11], aprendizado de vocabulário [21] e compreensão de uma língua estrangeira [14], para citar alguns.

As estratégias de criação de legendas se situam entre dois extremos. De um lado, legendas podem ser geradas e sincronizadas automaticamente com o conteúdo audiovisual usando reconhecimento automático de fala [5]. Do outro, a criação de legendas pode ser realizada com um esforço considerável, manualmente, por pessoas dedicadas a esse trabalho. Se profissionais podem ser contratados para criar e sincronizar legendas na fase de pós-produção do conteúdo audiovisual, entusiastas também podem se voluntariar para fazer este tipo de trabalho, como tem ocorrido na Web no caso de séries e filmes populares compartilhados *on-line*.

Quanto ao processo de distribuição, as legendas podem ser gravadas de forma definitiva sobre o conteúdo audiovisual, ou encapsuladas e distribuídas juntamente com DVDs¹, Blu-Rays, conteúdos televisivos e serviços de vídeo sob demanda, como o Netflix, HBO GO e Amazon Prime. Já no caso da geração por entusiastas, as legendas resultantes (normalmente, arquivos textuais no formato *SubRip Text* ou SRT) são compartilhadas *on-line* separadamente do conteúdo audiovisual.

Além das questões inerentes de autorização e direitos autorais dos conteúdos anotados, que vão além do escopo deste trabalho, a atividade de *legendadores voluntários* possui características próprias e pouco reportadas na literatura. Isso a difere fundamentalmente da atividade profissional tanto do ponto de vista de criação, quanto de compartilhamento de legendas. Nesse contexto, a contribuição deste trabalho está associada à investigação das questões relacionadas às práticas colaborativas do processo realizado por entusiastas. Baseado em resultados preliminares obtidos através de entrevistas com legendadores voluntários, formalizamos o processo de criação e compartilhamento de legendas realizado por esses entusiastas. Para isso, utilizamos um modelo de gerenciamento de processos de negócios e o comparamos a modelos *crowdsourcing* e de colaboração entre voluntários anônimos (por ex., Wikipedia), destacando desafios e oportunidades relevantes no contexto de aplicações e sistemas multimídia.

2 O PROCESSO DE LEGENDAGEM

Na literatura científica, são escassas as referências ao processo de criação e compartilhamento na Web de legendas para conteúdo

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

WebMedia '17, October 17–20, 2017, Gramado, Brazil.

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5096-9/17/10...\$15.00

<https://doi.org/10.1145/3126858.3131592>

¹ Algumas tecnologias mencionadas, se desconhecidas, podem ser facilmente identificadas através de uma pesquisa *on-line*; Portanto, elas não serão referenciadas.

audiovisual de terceiros; então, a metodologia que utilizamos nessa investigação foi essencialmente exploratória. Nesse processo, contamos com a ajuda de um legendador voluntário ao qual temos acesso. As informações apresentadas neste trabalho foram inicialmente elicitadas a partir da interação com esse entusiasta. Posteriormente, essa pessoa nos auxiliou no recrutamento de 4 legendadores para participar de entrevistas individuais. O critério de recrutamento foi de que cada um dos entrevistados desempenhasse um papel distinto no processo de legendagem². As entrevistas duraram em média 30min e seguiram um formato semiestruturado, onde foi incentivada a utilização do protocolo *think aloud*. Os participantes não receberam compensação e tivemos o consentimento para que o áudio das entrevistas fosse gravado.

Com base em discussões e reflexões sobre o material coletado, nesta seção apresentamos uma visão geral do processo de legendagem. Nossas observações são apresentadas de forma agregada e não identificadas individualmente.

2.1 Equipes de Legenda

A partir das entrevistas, identificamos que, no Brasil, o site Legendas.tv (LTV - <http://legendas.tv>) representa a principal comunidade de entusiastas, cuja interação e colaboração têm como resultado a produção de conteúdo, no caso, legendas para séries e filmes. De acordo com nossos entrevistados, o LTV funciona não só como o ponto focal de compartilhamento de legendas, mas também como gestor das séries e filmes de interesse que precisam ser legendados. Para viabilizar a criação de legendas para os conteúdos, o LTV conta com a colaboração de grupos de entusiastas que se auto organizam em equipes. O LTV coordena a distribuição de projetos (termo usado para descrever a atividade associada à criação de legenda para um filme ou de um episódio específico) entre equipes para evitar, por exemplo, que sejam produzidas mais de uma legenda para um mesmo conteúdo audiovisual.

Os projetos de legendagem são distribuídos de acordo com os perfis de trabalho da cada equipe. Por exemplo, se uma equipe costuma legendar projetos do gênero comédia, um novo projeto de comédia provavelmente será indicado para ela. Descobrimos ainda que, em algumas situações, como no caso dos *blockbusters* ansiosamente esperados pelos fãs usuários do LTV e com requisitos de tempo mais rígidos para a entrega, pode acontecer de mais de uma equipe legendar o mesmo projeto.

2.2 Papéis e Responsabilidades

Segundo os entrevistados, um projeto típico de criação de legendas considera que cada membro de uma equipe desempenha um papel associado a um conjunto de atividades e responsabilidades que devem ser cumpridas para que um projeto tenha êxito. Nesse estudo, descobrimos que, geralmente, são 4 os papéis que um membro de equipe pode assumir. São eles: *Legender* (termo usado para caracterizar a função de quem cria legendas), Revisor 1, Revisor 2 e Moderador.

O *Legender* é o papel mais básico do processo de legendagem. Ele normalmente escolhe seu pacote de trabalho ou *slot* (normalmente, de 5min ou 120 linhas, quando a versão *closed caption* está disponível) para traduzir e sincronizar dentre os projetos disponíveis, e terminado o trabalho, envia suas contribuições ao Revisor 1.

O Revisor 1, papel associado ao segundo nível hierárquico, tem como principal responsabilidade a revisão dos *slots* enviados pelos *Legenders*. Eventualmente, ele precisa corrigir os erros de sincronia/tradução, uso incorreto dos padrões e erros de *spotting* (Seção 2.3). O Revisor 1 fica ainda encarregado de enviar os *slots* revisados ao Revisor 2 e de dar *feedback* para o *Legender*, informando erros e problemas encontrados no *slot* recebido.

Já o Revisor 2 é responsável por dividir o projeto em *slots*, abrir uma chamada de participação aos colaboradores (normalmente por e-mail), revisar os *slots* recebidos e enviar *feedback* ao Revisor 1. Ele também fica encarregado de ajustar a sincronia da legenda para as diferentes distribuições de vídeo e publicar a legenda final no site do LTV (em algumas equipes, a divisão de *slots* e publicação podem ser feitas pelo Moderador).

O Moderador, papel associado à gerência da equipe, escolhe em comum acordo com o administrador do LTV os filmes e séries para os quais a equipe criará legendas. Ele também define os revisores de um projeto, intermedia a comunicação entre *Legenders* e Revisores, responde e-mails direcionados a equipe e coordena treinamentos de *Legenders* iniciantes.

De acordo com os resultados das participações em projetos, colaboradores podem ser requisitados a desempenhar outro papel. Assim, todo moderador possui os conhecimentos de Revisor 2, de Revisor 1 e de *Legender*. Por fim, os entrevistados evidenciaram que um mesmo colaborador pode executar diferentes papéis em um mesmo ou em diferentes projetos.

2.3 Padrões de Legendagem

De acordo com nossos quatro entrevistados, para se tornar um *Legender*, uma pessoa precisa passar por um treinamento para aprender noções de como fazer a tradução e a sincronia das legendas dentro dos padrões estabelecidos pela equipe. Após o treinamento o candidato se torna membro da equipe e tem autorização para participar de projetos. O responsável por treinar os candidatos a *Legender* normalmente é o Revisor 2.

Os entrevistados disseram que o *Legender* em treinamento recebe uma cartilha ou arquivo contendo informações sobre os padrões de exposição mínima e máxima da legenda na tela, a quantidade de caracteres por linha (CPL) e por segundo (CPS). Esses parâmetros podem variar ligeiramente entre equipes, mas no geral, a exposição mínima fica em torno de 1,3s e máxima de 5,0s. Já o CPL aceitável fica em torno de 32 caracteres e o CPS em torno de 20 caracteres. Esses padrões buscam assegurar um tempo suficiente para a compreensão do texto pelo espectador. Entretanto, os entrevistados não souberam dizer se esses padrões são os mesmos da indústria cinematográfica. Outros padrões ensinados são relacionados ao uso do travessão, itálico, reticências e interjeições.

Além disso, os entrevistados disseram que existem regras referentes à tradução, sincronia e finalização da legenda. No caso

² Os termos 'legendar' e 'legendagem' são utilizados por legendadores para se referir à atividade de criação de legendas que envolve as fases de tradução e sincronização.

Tabela 1: Exemplos de treinamento de *spotting*

Spotting Ruim	Spotting Bom
Precisa nos explicar o que o faz especial.	Precisa nos explicar o que o faz especial.
Não posso ir a Las Vegas nos fins de semana.	Não posso ir a Las Vegas nos fins de semana.
Em todo caso, contar cartas não é ilegal.	Em todo caso, contar cartas não é ilegal.

da tradução, os *Legenders* recebem instruções sobre como traduzir gírias, onde encontrar os nomes de personagens, quando usar ou não pronomes, e como traduzir textos e letreros que aparecem na tela. Ainda no que se refere à tradução, um conceito muito importante para a fluidez da legenda é o chamado *spotting*. Em linhas gerais, um *spotting* bom visa minimizar o efeito das quebras de linha na compreensão de uma frase (no geral, se uma linha ultrapassa 27 caracteres, sugere-se que uma quebra de linha seja inserida). A Tabela 1 apresenta exemplos de *spotting* extraídos de uma cartilha de treinamento.

No caso da sincronia, foi apontado que o ideal é que uma legenda comece cerca de 200ms antes da fala e se encerre após 400ms. As instruções de treinamento enfatizam ainda a necessidade de que sejam usados o corretor ortográfico e a busca por erros antes da finalização da tarefa (e sua consequente entrega). De acordo com nossos entrevistados, essas funcionalidades são geralmente integradas às ferramentas de edição utilizadas pelos *Legenders*.

3 MODELAGEM CONCEITUAL

A Figura 1 ilustra a modelagem conceitual do processo de criação e compartilhamento de legendas usando o padrão BPMN (*Business Process Model and Notation*). Optamos por utilizar esse padrão pois o mesmo facilita o entendimento dos procedimentos de negócio através de uma notação gráfica. As raias horizontais representam os papéis interpretados pelos colaboradores no processo de legendagem e suas atividades associadas. Este diagrama foi feito de maneira incremental, sendo refinado e, posteriormente, validado por todos os nossos entrevistados.

Como discutido na seção anterior, após a moderação escolher um filme ou série e os revisores do projeto, o Revisor 2 divide os *slots* (baseado no tempo ou número de linhas do *closed caption*) e abre a chamada convidando os membros da equipe a participarem do projeto. Dependendo da equipe, a distribuição de *slots* e todas as informações referentes ao projeto estarão em uma planilha compartilhada do *Google Sheets* ou no site da equipe, em uma área restrita aos membros. As informações de um projeto vão desde o *release* de trabalho (distribuição do vídeo) e legenda base (quando disponível), até os termos específicos daquele projeto e informações de contato dos revisores.

O processo de legendagem completo para um episódio de série de TV dura, em média, sete dias. Inicialmente, *Legenders* têm de 2 a 3 dias para legendar o *slot* escolhido e enviar ao Revisor 1, que por sua vez terá 2 dias para revisar os *slots*

recebidos (a média gira em torno de 7 *slots* por Revisor 1) e enviar ao Revisor 2. O Revisor 2 também terá 2 dias para revisar os *slots* recebidos, enviar *feedback* ao Revisor 1 (e a moderação sobre a qualidade do trabalho do Revisor 1) e fazer ajustes necessários para os *releases*, para que a legenda possa ser publicada no site do LTV.

4 DISCUSSÃO

Na literatura podemos encontrar vários trabalhos com enfoque em problemas inerentes à criação [4][9] e ao compartilhamento maciço de arquivos de legendas em bases de dados *on-line* [8][18][19]. No contexto do WebMedia, também é possível listar inúmeros esforços dedicados à anotação de mídia para enriquecimento de conteúdo audiovisual [10][15][17]. Contudo, pouco se sabe sobre as práticas colaborativas de anotação para enriquecimento de conteúdo audiovisual, em particular, no contexto de legendagem para vídeos de séries de TV e filmes.

Mais próximo ao nosso trabalho, Peloi [13] faz uma reflexão das práticas de confecção e distribuição gratuita de cópias de legendas de filmes, séries etc., sob uma perspectiva de cidadania e resistência. Apesar de nossa pesquisa ter pontos de similaridade, nós focamos em um problema de pesquisa complementar: Compreender e descrever como se organizam e como trabalham grupos de entusiastas que criam e compartilham na Internet legendas para conteúdo audiovisual de terceiros.

O diagrama BPMN da Figura 1 alude ao modelo 3C de colaboração cujas dimensões são comunicação, coordenação e cooperação [7]. Além da coordenação, a comunicação e a cooperação são fatores extremamente importantes para criação de legendas. Por exemplo, dúvidas que surgem enquanto um *Legender* está fazendo um *slot* muitas vezes são esclarecidas através de grupos do WhatsApp, planilhas da equipe ou e-mails.

A partir desta observação, pode-se afirmar que o processo de legendagem estudado não está alinhado com o conceito de *crowdsourcing*, uma vez que os *Legenders* se comunicam frequentemente durante a realização das tarefas. A comunicação entre executores de tarefa viola um dos princípios básicos das execuções individuais e isoladas das tarefas para a composição final de um resultado no *crowdsourcing* tradicional.

Por outro lado, percebemos que esse processo se assemelha ao processo colaborativo utilizado na enciclopédia *on-line* Wikipedia, onde usuários anônimos se auto organizam para alcançar um objetivo. Além disso, podemos notar semelhanças com o processo de revisão por pares e avaliação de contribuições. Por fim, vale destacar que a comunidade do Wikipedia também desenvolveu um conjunto de políticas e recomendações que servem de apoio para as atividades de seus colaboradores.

5 CONCLUSÃO

Neste artigo, apresentamos nossas primeiras observações sobre o processo de legendagem para vídeos compartilhados *on-line*. A principal contribuição do nosso trabalho está na identificação dos diferentes papéis existentes em comunidades de legendadores voluntários e, na consequente, formalização das

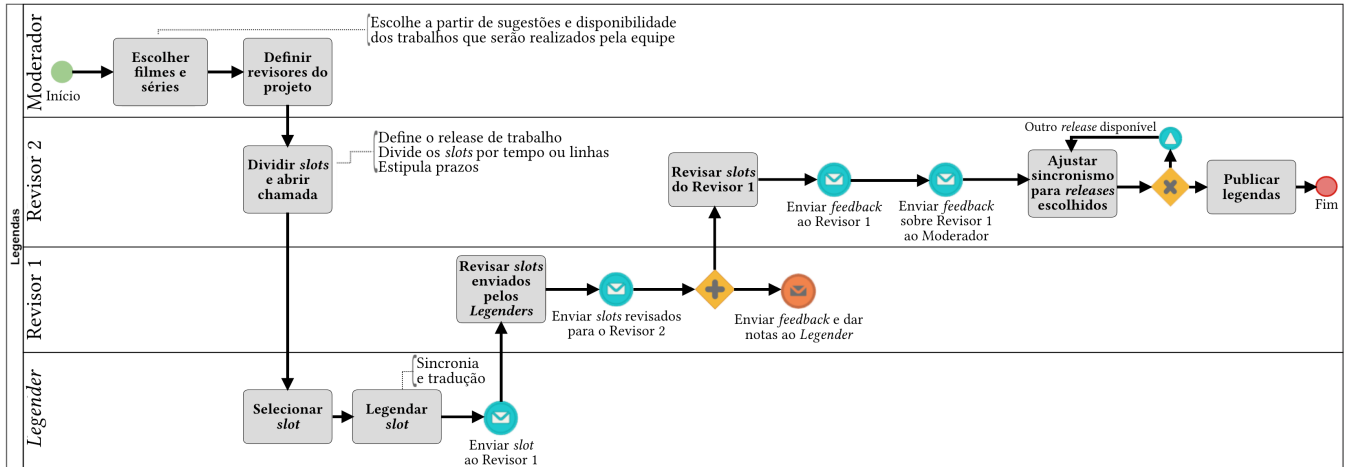


Figura 1. Diagrama BPMN do processo de legendagem na Web.

principais práticas e atividades que caracterizam o processo colaborativo de criação e compartilhamento de legendas. Nossos resultados preliminares, obtidos a partir de entrevistas individuais com entusiastas, indicam que este processo, apesar de voluntário, segue uma hierarquia e fluxo de trabalho rigorosos, exigindo disciplina e comprometimento dos participantes.

Com relação a trabalhos futuros, pretendemos investigar as razões, motivações e objetivos que levam legendadores voluntários a dedicar tempo e energia a um processo sem remuneração como esse. Além disso, outras direções de investigação que estamos avaliando para dar continuidade a esse trabalho são a consideração da abordagem de *crowdsourcing* e automatização de partes do processo de legendagem.

AGRADECIMENTOS

Os autores agradecem aos *Legenders* pela participação nos estudos que permitiram a realização deste trabalho. Jessica O. Brito é bolsista da FAPES (#789/2015) e Celso A. S. Santos do CNPq (#312148/2014-3) e FAPES (#67927378/2015).

REFERÊNCIAS

- [1] Dick C. A. Bulterman, Pablo Cesar, and Rodrigo L. Guimarães. 2013. Socially-aware multimedia authoring: Past, present, and future. *ACM Trans. Multimedia Comput. Commun. Appl.* 9, 1s, 35 (October 2013), 23 pages.
- [2] Bruna C.R. Cunha, Olibario J.M. Neto, and Maria G.C. Pimentel. 2013. A heuristic evaluation of a mobile annotation tool. In *Proc. of WebMedia*, 89-92.
- [3] Leonardo A. Domingues, Virginia P. Campos, Tiago M. U. Araújo, and Guido L. S. Filho. 2016. Accessibility in Digital Cinema: A Proposal for Generation and Distribution of Audio Description. In *Proc. of Webmedia*, 119-126.
- [4] Roberto Fagã Jr., Vivian G. Motti, Renan G. Cattelan, Cesar A. C. Teixeira, and Maria G. C. Pimentel. 2010. A social approach to authoring media annotations. In *Proc. of the ACM DocEng*, 17-26.
- [5] Maria Federico and Marco Furini. 2014. An automatic caption alignment mechanism for off-the-shelf speech recognition technologies. *Multimedia Tools Appl.* 72, 1 (Sep. 2014), 21-40.

- [6] José L. M. Ferriol. 2007. An empirical and descriptive study of the translation method for dubbing and subtitling. *A Tool for Social Integration? Audiovisual Translation from Different Angles*, No 6 (2007), 171-184.
- [7] Hugo Fuks, Alberto B. Raposo e Marco A. Gerosa. 2003. Do modelo de colaboração 3c à engenharia de groupware. In *Proc. of WebMedia*, 0-8.
- [8] Rodrigo L. Guimarães, Priscilla Avegliano, and Lucas C. Villa Real. 2016. A Lightweight and Efficient Mechanism for Fixing the Synchronization of Misaligned Subtitle Documents. In *Proc. of the ACM DocEng*, 175-184.
- [9] Rodrigo L. Guimarães, Pablo Cesar, and Dick C. A. Bulterman. 2012. Let me comment on your video: supporting personalized end-user comments within third-party online videos. In *Proc. of WebMedia*, 253-260.
- [10] Fons Kuijk, Rodrigo L. Guimarães, Pablo Cesar, and Dick C. A. Bulterman. 2009. Adding dynamic visual manipulations to declarative multimedia documents. In *Proc. of the ACM DocEng*, 149-152.
- [11] Raja S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. 2013. Captions versus transcripts for online video content. In *Proc. of W4A*, 32, 4 pages.
- [12] Manuella A. C. B. Lima, Tiago M.U. de Araújo, and Erickson S. de Oliveira. 2015. Incorporation of Syntactic-Semantic Aspects in a LIBRAS Machine Translation Service to Multimedia Platforms. In *Proc. of WebMedia*, 133-140.
- [13] Danilo Peloi. 2012. Legendamento Não Autorizado na Rede: Práticas de Cópia e Resistência. In: *Anais do Congresso Int. Interdisciplinar Em Sociais e Humanidades*, 1, 2012, Niterói. ISSN 2316-266X
- [14] K. Rooney. 2014. The Impact of Keyword Caption Ratio on Foreign Language Listening Comprehension. *Int. Journal of Computer-Assisted Language Learning and Teaching*. 4, 2 (Apr. 2014), 11-28.
- [15] Celso A. S. Santos, Alexandre V. S. G. Santos, and Tatiana A. Tavares. 2007. Uma Estratégia para a Construção de Ambientes para a Descrição Semântica de Vídeos. In *Proc. of WebMedia*, 274-281.
- [16] Ricardo Segundo, Marcello N. Amorim e Celso A. S. Santos. 2016. Crowdsourcing & Multimedia: Enhancing Multimedia Activities with the Power of Crowds. In: *Proc. of WebMedia*, 11-12.
- [17] Marcelo F. Sousa, Raoni Kulesza, and Carlos A. Ferraz. 2016. A Model-driven Approach for MulSeMedia Application Domain. In *Proc. of Webmedia*, 111-118.
- [18] Jörg Tiedemann. 2008. Synchronizing translated movie subtitles. In *Proc. of the Int. Conf. on Language Resources and Evaluation (LREC '08)*, 5 pages.
- [19] Lucas C. Villa Real, Rodrigo L. Guimarães, and Priscilla Avegliano. 2015. Dynamic Adjustment of Subtitles Using Audio Fingerprints. In *Proc. of the ACM international conference on Multimedia (MM '15)*, 975-978.
- [20] Fangzhou Wang, Hidehisa Nagano, Kunio Kashino, and Takeo Igarashi. 2017. Visualizing Video Sounds With Sound Word Animation to Enrich User Experience. *Trans. Multi.* 19, 2 (February 2017), 418-429.
- [21] Yeshuang Zhu, Yuntao Wang, Chun Yu, Shaoyun Shi, Yankai Zhang, Shuang He, Peijun Zhao, Xiaojuan Ma, and Yuanchun Shi. 2017. ViVo: Video-Augmented Dictionary for Vocabulary Learning. In *Proc. CHI*, 5568-5579.