

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**C2C: UM CHAT BILÍNGUE COM APOIO DE SENSO
COMUM**

BRUNO AKIO SUGIYAMA

ORIENTADORA: PROF. DR. JUNIA COUTINHO ANACLETO

COORIENTADORA: PROF. DR. HELENA DE MEDEIROS CASELI

São Carlos - SP
Setembro/2011

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**C2C: UM CHAT BILÍNGUE COM APOIO DE SENSO
COMUM**

BRUNO AKIO SUGIYAMA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Interação Humano-Computador.
Orientadora: Dr. Junia Coutinho Anacleto

São Carlos - SP
Setembro/2011

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S947cc

Sugiyama, Bruno Akio.

C2C : um chat bilíngue com apoio de senso comum /
Bruno Akio Sugiyama. -- São Carlos : UFSCar, 2011.
83 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2011.

1. Ciência da computação. 2. Interfaces de usuário
(Sistema de computador). 3. Interface e comunicação. 4.
Linguagem - tradução automática. 5. OMCS (Open Mind
Common Sense). 6. Base de conhecimento. I. Título.

CDD: 004 (20^a)

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

*“C2C: Um chat bilíngüe com
apoio de senso comum”*

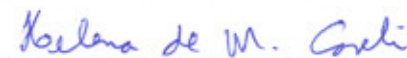
BRUNO AKIO SUGIYAMA

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Ciência da
Computação da Universidade Federal de São
Carlos, como parte dos requisitos para a
obtenção do título de Mestre em Ciência da
Computação

Membros da Banca:



Profa. Dra. Junia Coutinho Anacleto
(Orientadora - DC/UFSCar)



Profa. Dra. Helena de Medeiros Caseli
(Co-Orientadora - DC/UFSCar)



Prof. Dr. Cesar Augusto Camillo Teixeira
(DC/UFSCar)



Profa. Dra. Carla Dal Sasso Freitas
(UFRGS)

São Carlos
Outubro/2011

Dedico este trabalho aos meu pais, Luiz e Terezinha.

AGRADECIMENTO

Agradeço a minha namorada por todo apoio durante o mestrado, aos meus pais por todo incentivo e apoio, aos colegas e ex-colegas do laboratório pela ajuda e companheirismo, a todos os voluntários do estudo piloto deste trabalho, a minha coorientadora por toda ajuda e disponibilidade e a minha orientadora por todas as oportunidades oferecidas.

Não é triste mudar de ideias, triste é não ter ideias para mudar.

Sarão de Stararé

RESUMO

Neste trabalho é descrito como o conhecimento de senso comum em conjunto com a tradução automática pode apoiar a comunicação entre pessoas de diferentes culturas. Para verificar a viabilidade do senso comum foi desenvolvido um *chat* bilíngue chamado Culture-to-Chat ou C2C, que, além de prover um canal de comunicação, possui mecanismos que auxiliam o usuário na criação de mensagens em língua estrangeira. No campo da comunicação mediada por computador, percebe-se que as pessoas estão cruzando as fronteiras geográficas e tendo oportunidades de troca de experiências entre diferentes culturas. Para alguns usuários, essa troca de informações é feita em uma língua não nativa, o que pode ser uma tarefa difícil para eles. Algumas ferramentas computacionais que apoiam a comunicação propõem o uso de tradução automática para trabalhar com usuários falantes de diferentes línguas. O C2C, além de adotar tal estratégia, adota a rede semântica de conhecimento cultural construída colaborativamente via Web do projeto Open Mind Common Sense no Brasil (OMCS-Br) para trabalhar com expressões culturais, ou seja, termos cujo significado depende da cultura do usuário. Utilizando uma abordagem centrada no usuário focando-se em prototipação, o desenvolvimento do C2C e suas funcionalidades são apresentados por meio de protótipos de diferentes níveis de fidelidade (baixa, média e alta). Com o intuito de observar o uso dessa ferramenta computacional e coletar opiniões de usuários, foi realizado um estudo piloto envolvendo usuários brasileiros e canadenses. Tal estudo mostrou possíveis melhorias pra a ferramenta e apontou indícios de que este *chat* contribui na comunicação entre pessoas de diferentes culturas.

Palavras-chave: conhecimento cultural, comunicação mediada por computador, tradução cultural, senso comum, tradução automática, OMCS-Br.

ABSTRACT

In this research, we describe how common sense knowledge with machine translation can help the communication among people with different cultural background. In order to evaluate this possibility, we developed a bilingual chat called Culture-to-Chat or C2C that provides a communication channel and has resources that help its user to create messages in a second language. In the computer-mediated communication field, it is possible to notice that people are crossing geographic borders and having opportunities to share experiences among different cultures. Sharing information in a non native language can be difficult to some users. Some computational tools that support communication uses machine translation to aid users that need to work with different language. C2C also uses this approach and adopts a semantic network of cultural knowledge, collaboratively built on the Web through the Open Mind Common Sense in Brazil project (OMCS-Br) to work with cultural expression, in other words, terms whose meaning depends on user's culture. Following a user-centered design approach that focuses on prototyping, we present the development of C2C passing by low, middle and high fidelity prototypes. In order to observe how this computational tool is used and collect the opinion of target users, we perform a pilot study involving Brazilian and Canadian users. This study showed some enhancements for the tool and pointed evidences that this chat contributes to the communication between people with different cultural backgrounds.

Keywords: cultural knowledge, computer-mediated communication, cultural translation, common sense, machine translation, OMCS-Br.

LISTA DE FIGURAS

Figura 2-1 <i>Chat</i> com EN2ZH (CHEN e RAMAN, 2008).....	21
Figura 2-2 Interface do Polyglot para idioma inglês (AIKEN e MINA, 2009)	22
Figura 2-3 Interface do AnnoChat (YAMASHITA e ISHIDA, 2006)	22
Figura 2-4 Arquitetura do projeto OMCS-Br	25
Figura 2-5 <i>Template</i> para coleta de gírias.....	26
Figura 2-6 <i>Template</i> para coleta de significado de gírias	27
Figura 3-1 Primeiro protótipo do chat	32
Figura 3-2 Segunda versão do protótipo, agora denominado 2-chat	33
Figura 3-3 Protótipo de média fidelidade para o <i>chat</i> , agora denominado C2C	34
Figura 3-4 Primeiro protótipo de alta-fidelidade do C2C	35
Figura 3-5 Entrando na sala de bate-papo do C2C.....	36
Figura 3-6 Tela inicial do C2C	37
Figura 3-7 Interface do C2C em inglês para usuários estrangeiros	38
Figura 3-8 Exemplo de interação do usuário brasileiro no C2C	38
Figura 3-9 Edição e envio da mensagem em inglês no C2C	39
Figura 3-10 Arquitetura do C2C	40
Figura 3-11 Fluxo da informação no envio da mensagem do usuário brasileiro para o usuário estrangeiro.....	41
Figura 4-1 Interface do C2C sem o recurso de tradução cultural utilizada na Etapa 2 do estudo piloto	48
Figura 4-2 Exemplo de alguns campos do cadastro do usuário.....	49
Figura 4-3 Exemplos de perguntas do questionário pré-sessão do usuário brasileiro	50
Figura 4-4 Exemplos de perguntas do questionário pós-sessão do usuário brasileiro	50
Figura 4-5 Exemplos de perguntas do questionário pré-sessão do usuário canadense.....	51
Figura 4-6 Exemplos de perguntas do questionário pós-sessão do usuário canadense.....	52
Figura 4-7 Menores médias de mensagens enviadas por minuto	65

LISTA DE TABELAS

Tabela 2-1 Comparação entre ferramentas com TA	23
Tabela 4-1 Informações do cadastro dos usuários brasileiros	53
Tabela 4-2 Perfil dos usuários brasileiros	53
Tabela 4-3 Informações do cadastro dos usuários canadenses	54
Tabela 4-4 Perfil dos usuários canadenses	54
Tabela 4-5 Informações da sessão de bate-papo	55
Tabela 4-6 Respostas do questionário pós-sessão dos brasileiros.....	56
Tabela 4-7 Respostas do questionário pós-sessão dos canadenses.....	58
Tabela 4-8 Comparação do nível de inglês dos brasileiros.....	59
Tabela 4-9 Informações do cadastro dos usuários brasileiros (segunda etapa)	60
Tabela 4-10 Perfil dos usuários brasileiros (segunda etapa)	60
Tabela 4-11 Perfil dos usuários canadenses (segunda etapa).....	61
Tabela 4-12 Informações das sessões de bate-papo (segunda etapa).....	61
Tabela 4-13 Respostas do questionário pós-sessão dos brasileiros.....	62
Tabela 4-14 Respostas do questionário pós-sessão canadense (segunda etapa) ...	63
Tabela 4-15 Comparação do nível de inglês dos brasileiros (segunda etapa)	64

LISTA DE ABREVIATURAS E SIGLAS

AJAX - *Asynchronous Javascript and XML*

C2C – Culture-to-Chat

CMC – Comunicação Mediada por Computador

ESL – *English as a Second Language*

HTML - *HyperText Markup Language*

HTTP - *Hypertext Transfer Protocol*

LIA – Laboratório de Interação Avançada

OMCS-Br – *Open Mind Common Sense* no Brasil

PLN – Processamento de Língua Natural

SMT – *Statistical Machine Translation*

TA – Tradução Automática

UCD – *User Centered Design*

XML - *eXtensible Markup Language*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	13
1.1 Contexto.....	13
1.2 Motivação e Objetivos.....	14
1.3 Metodologia de Desenvolvimento do Trabalho.....	15
1.4 Organização do Trabalho.....	16
CAPÍTULO 2 - COMUNICAÇÃO MEDIADA POR COMPUTADOR.....	17
2.1 Considerações Iniciais.....	17
2.2 Comunicação mediada por computador.....	17
2.2.1 Os meios de comunicação providos pelo computador.....	18
2.2.2 A Internet e a Quebra de Fronteiras.....	19
2.3 Tradução Automática.....	19
2.4 Ferramentas Computacionais de Apoio a Comunicação.....	20
2.5 Senso Comum.....	24
2.5.1 O Projeto OMCS-Br.....	24
2.6 Considerações Finais.....	29
CAPÍTULO 3 - CULTURE-TO-CHAT (C2C).....	30
3.1 Considerações Iniciais.....	30
3.2 Modelo de Desenvolvimento Baseado em Prototipação.....	30
3.3 A Arquitetura do C2C.....	40
3.4 O Modelo de Fluxo da Informação do C2C.....	41
3.5 O Processo de Retroalimentação da Base de Conhecimento do OMCS-Br.....	42
3.6 Considerações Finais.....	43
CAPÍTULO 4 - ESTUDO PILOTO.....	44
4.1 Considerações Iniciais.....	44
4.2 Local de Estudo.....	44
4.3 Planejamento.....	45
4.4 Etapas do Estudo Piloto.....	47
4.5 Preparação dos Questionários Utilizados no Estudo Piloto.....	48

4.5.1	Questionário Pré-Sessão: Usuário Brasileiro	48
4.5.2	Questionário Pós-Sessão: Usuário Brasileiro.....	50
4.5.3	Questionário Pré-Sessão: Usuário Canadense	51
4.5.4	Questionário Pós-Sessão: Usuário Canadense	52
4.6	Primeira Etapa do Estudo Piloto.....	52
4.6.1	Respostas do Questionário Pré-Sessão: Usuário Brasileiro	53
4.6.2	Respostas do Questionário Pré-Sessão: Usuário Canadense	54
4.6.3	Interação dos Usuários com C2C com Apoio da Tradução Cultural	55
4.6.4	Respostas do Questionário Pós-Sessão: Usuário Brasileiro.....	56
4.6.5	Respostas do Questionário Pós-Sessão: Usuário Canadense	58
4.7	Segunda Etapa do Estudo Piloto.....	59
4.7.1	Respostas do Questionário Pré-Sessão: Usuário Brasileiro	60
4.7.2	Respostas do Questionário Pré-Sessão: Usuário Canadense	61
4.7.3	Utilizando o C2C sem Apoio de Tradução Cultural	61
4.7.4	Respostas do Questionário Pós-Sessão: Usuário Brasileiro.....	62
4.7.5	Respostas do Questionário Pós-Sessão: Usuário Canadense	63
4.8	Análise Geral dos Resultados	64
4.9	Considerações Finais	65
	CAPÍTULO 5 - CONCLUSÕES E TRABALHOS FUTUROS.....	66
5.1	Síntese dos Principais Resultados	66
5.2	Trabalhos Futuros	67
	REFERÊNCIAS.....	70
	APÊNDICE A.....	75
	APÊNDICE B	77
	APÊNDICE C	79
	APÊNDICE D.....	81
	APÊNDICE E	83

Capítulo 1

INTRODUÇÃO

1.1 Contexto

A Internet teve um dos crescimentos mais rápidos da história em comparação com outras tecnologias (LEINER et al, 2009). Ela tem transformado substancialmente a comunicação do dia a dia nos campos profissional, educacional e pessoal (CASTELLS, 2005). Baseando-se em McQuail (2010), a comunicação mediada por computador (CMC) pode ser definida como qualquer interação que visa a comunicação por meio de dois ou mais computadores conectados.

Alguns exemplos de CMC são *Instant Messaging*, e-mail, salas de bate-papo ou *chat*, fóruns e mensagens de texto (SMS). Segundo Warschauer (1997), essas formas de comunicação por longa distância dão abertura a uma série de possibilidades interculturais (em inglês, *cross-cultural*). Em geral, a comunicação entre indivíduos de diferentes culturas, que é chamada de *Cross-Cultural Communication*, apresenta mais dificuldade e demanda maior esforço por parte de seus participantes (KAI, 2005).

Cultura pode ser definida como um conjunto complexo que inclui os conhecimentos, crenças, artes, moral, leis e qualquer outra habilidade ou capacidade adquiridos por uma pessoa como membro da sociedade (TYLOR, 1871). Entende-se que ela representa a nacionalidade, etnia ou região de um grupo de indivíduos.

Os *softwares* sensíveis à cultura do usuário (ANACLETO, 2006) usam o conhecimento cultural para prover maior usabilidade, facilitando a realização de

determinada tarefa do usuário. O Projeto *Open Mind Common Sense* no Brasil (OMCS-Br)¹, criado em 2005, visa coletar o senso comum da população brasileira para tornar as aplicações computacionais mais contextualizadas. O senso comum é um tipo de conhecimento cultural que pode ser definido como o conhecimento do dia a dia das coisas baseadas nas crenças e experiências da vida considerando-se o tempo e o espaço.

1.2 Motivação e Objetivos

Existe um grande número de pessoas que se comunicam utilizando um computador. No Brasil, por exemplo, mais de 22 milhões de internautas utilizam alguma ferramenta computacional para se comunicar (LIPSMAN, 2010). Alguns exemplos dessas ferramentas são o ICQ, MSN, Windows Live Messenger, Google Talk, Skype.

Apesar de proporcionarem um canal de comunicação entre pessoas que se encontram em localidades diferentes, poucas ferramentas tratam da comunicação entre pessoas que falam diferentes línguas e possuem diferentes culturas (AIKEN e PARK, 2009). Uma abordagem adotada para auxiliar o usuário com a língua estrangeira é o uso da tradução automática. Houve grandes avanços na área (CASELI e NUNES, 2009), mas a tradução automática ainda não cobre todas as necessidades dos usuários, ou seja, nem sempre o texto traduzido possui coerência e o mesmo sentido do texto original (CHEN e RAMAN, 2008; AIKEN e PARK, 2009).

Uma abordagem utilizada por alguns projetos (ANACLETO et al., 2006; SILVA e ANACLETO, 2009) é o uso de senso comum, que é uma forma de conhecimento cultural, para desenvolver ferramentas contextualizadas que consideram a cultura de seus usuários. Tal conhecimento é utilizado como um recurso adicional das ferramentas para auxiliar o usuário na realização de determinada tarefa.

Um objetivo deste trabalho é investigar o uso do senso comum em conjunto com a tradução automática de maneira a ajudar a comunicação de usuários falantes de diferentes línguas e de diferentes culturais. Desse modo, é apresentado o

¹ <http://www.sensocomum.ufscar.br>

desenvolvimento de uma ferramenta de bate-papo (*chat*) chamada de Culture-to-Chat ou C2C que auxilia os usuários na criação de mensagens em língua estrangeira.

O C2C tem como objetivos prover um canal de comunicação entre pessoas, auxiliar o usuário na escrita de mensagens em língua estrangeira e na comunicação com pessoas de outras culturas. Para isso, ele conta com recursos que utiliza o conhecimento de senso comum do projeto *Open Mind Common Sense* no Brasil (OMCS-Br) e modelos de tradução automática estatística (método considerado o estado da arte), permitindo a elaboração de traduções contextualizadas culturalmente.

1.3 Metodologia de Desenvolvimento do Trabalho

O presente trabalho foi desenvolvido seguindo a abordagem *User-Centered Design* (UCD) ou Projeto Centrado no Usuário (NORMAN e DRAPER, 1986), baseado na proposta apresentada em (ANACLETO et al., 2010) tendo o foco na prototipação. O ciclo de vida do projeto é composto por iterações ou fases. A cada iteração são levantadas as seguintes informações: (1) O objetivo da fase; (2) A estratégia de design adotada nesta iteração; (3) As questões que surgiram durante a iteração; (4) o protótipo utilizado para responder as questões; (5) os *stakeholders* envolvidos nesta etapa; (6) as respostas das questões levantadas.

Para a avaliação dos resultados do trabalho, foi adotada a estratégia de estudo piloto, também conhecido como experimento piloto. O estudo piloto é um estudo preliminar de menor escala, conduzido antes da pesquisa principal para verificar a viabilidade ou melhorar o *design* da pesquisa. Segundo Nielsen (2011), o estudo piloto possui alguns critérios mais flexíveis em comparação ao estudo de caso, como por exemplo, com relação à escolha dos usuários participantes na avaliação da ferramenta.

O estudo piloto pode incluir tanto evidências qualitativas como evidências quantitativas. A análise qualitativa procura responder as questões de pesquisas por meio de organização, interpretação e categorização dos dados, com finalidade de adquirir conhecimento e dar significado a uma determinada experiência. Ambas as análises são interessantes para este trabalho, visto que a análise quantitativa pode

informar dados com relação à interação com a ferramenta proposta e a análise qualitativa ajuda na explicação dos dados quantitativos.

1.4 Organização do Trabalho

Além deste Capítulo introdutório, este trabalho encontra-se organizado em outros quatro Capítulos: no Capítulo 2, são descritos a área de comunicação mediada por computador, assim como alguns exemplos de ferramentas que permitem a comunicação entre pessoas e recursos envolvidos nesse processo; no Capítulo 3 é apresentado o processo de desenvolvimento e as funcionalidades do C2C; no Capítulo 4 é descrito o estudo piloto realizado com o intuito de avaliar o uso do C2C e coletar a opinião dos usuários sobre o mesmo; por fim, no Capítulo 5 é apresentada a conclusão deste trabalho e os possíveis trabalhos futuros.

Capítulo 2

COMUNICAÇÃO MEDIADA POR COMPUTADOR

2.1 Considerações Iniciais

Este capítulo descreve a forma com que o computador tem sido utilizado na comunicação interpessoal, assim como seu impacto neste tipo de comunicação. Alguns trabalhos relacionados e tecnologias referentes a este trabalho também são apresentados.

Este capítulo está dividido nas seguintes seções: 2.2 apresenta o campo da comunicação mediada por computador; 2.3 descreve o recurso de tradução automática; 2.4 aborda algumas ferramentas de apoio à comunicação que utilizam tradução automática; 2.5 descreve um tipo de conhecimento cultural: o senso comum e; 2.6, as considerações finais.

2.2 Comunicação mediada por computador

De acordo com December (1996), a comunicação mediada por computador (CMC) pode ser definida como um processo pelo qual as pessoas criam, trocam e percebem a informação usando sistemas de telecomunicação por rede que facilitam a codificação, transmissão e decodificação de mensagens. Pesquisas na área de

CMC englobam uma série de perspectivas interdisciplinares, combinando pessoas, tecnologia, processos e efeitos.

Existem duas possíveis classificações para a CMC: comunicação síncrona (tempo real) e assíncrona (tempo de espera). A comunicação síncrona ou em tempo real, por exemplo, a conversa face-a-face (f2f), conversa pelo telefone ou até palestras (um-para-muitos) também tem seu equivalente na CMC por meio de salas de bate-papo e ambientes similares. Na comunicação assíncrona, existe um tempo de demora entre o envio da mensagem e leitura da mesma, como por exemplo, a escrita de cartas ou envio de faxes. Os equivalentes *online* dessa forma de comunicação são o e-mail, as listas de discussão (fóruns) e as conferências por computador.

A CMC provê um processo complexo de interação entre os participantes (ROMISZOWSKI e MASON, 2004). Ela combina a natureza da comunicação escrita com a velocidade e dinamismo da comunicação falada. Recentemente, outras tecnologias utilizam de vídeo, som e outras compilações de media para agregar tais recursos na comunicação. Praticamente, não existem limites para as possibilidades de interação e *feedback*, onde os usuários utilizam a criatividade e a imaginação em seu envolvimento pessoal.

2.2.1 Os meios de comunicação providos pelo computador

O computador oferece diversos meios para a comunicação interpessoal. Como exemplo de meio assíncrono, o email é uma aplicação computacional de bastante sucesso até hoje (WHITTAKER, 2005; MAHUD et al, 2011), sendo uma forma de trocar informações por meio da qual mensagens são enviadas de um computador para o outro por meio de um sistema de telecomunicações. Ele suporta comunicação interativa entre muitos participantes. Recentemente, as redes sociais como facebook² e twitter³ também promovem a comunicação assíncrona.

Dos meios síncronos, existem as possibilidades de comunicação via textual (por exemplo, *instant messenger*) e via oral (por exemplo, Skype). Em particular, o bate-papo ou *chat* é considerado ao mesmo tempo uma plataforma interativa e um meio que potencializa o aprendizado significativo (TENG, 2010). Em particular, o

² <http://www.facebook.com>

³ <http://twitter.com>

chat permite uma troca dinâmica virtual de informação em tempo real onde os participantes leem e respondem a mensagens imediatamente. Este tipo de ferramenta é bastante usado por estudantes que estão em processo de aprendizagem de uma segunda língua (LAM, 2004; HIRVELA, 2006), visto que oferece oportunidades reais de comunicação, aumentando a taxa de troca de informação e a quantidade de interação (PARAMSKAS, 1999).

2.2.2 A Internet e a Quebra de Fronteiras

A Internet propulsionou as tecnologias da CMC, permitindo que a comunicação interpessoal fosse realizada por meio de diversas configurações (GARCIA et al, 2009). Ela proveu a oportunidade para que as fronteiras da comunicação entre indivíduos de localidades e culturas diferentes fossem cruzadas (LEE e LEE, 2010).

Com essa quebra de fronteira, ocorre a conversa envolvendo pessoas que falam diferentes línguas. A comunicação entre pessoas que falam diferentes línguas, em geral, é realizada em um dos idiomas (YAMASHITA e ISHIDA, 2006), sendo que é necessário que um dos participantes utilize uma língua não nativa (língua estrangeira ou segunda língua) para se comunicar. Para os participantes, este tipo de comunicação se torna difícil (KIM e BONK, 2002) tanto por causa da barreira da língua quanto por causa das diferenças culturais.

Uma abordagem para tratar este problema é a utilização de recursos de tradução automática, descrita na próxima seção e de senso comum, descrito na Seção 2.5.

2.3 Tradução Automática

A Tradução Automática (TA) é uma das mais antigas e importantes áreas de pesquisa do Processamento de Língua Natural (PLN). É um tema cada vez mais atual no cenário de globalização no qual a disseminação de informações multilíngues é cada vez maior e aponta para a necessidade de traduções rápidas, eficientes e de baixo custo.

A tarefa de TA consiste em, dado um texto escrito em uma língua de origem (texto-fonte), produzir uma versão do mesmo texto em outra língua natural (texto-alvo). O grande problema desse processo, segundo Nirenburg (1987), é gerar, como resultado, uma versão (texto-alvo) que mantenha o significado mais próximo possível daquele originalmente existente no texto-fonte. Para alguns autores, inclusive, a tradução humana é considerada uma arte (HUTCHINS, 1998) já que envolve escolhas pessoais, não sendo simplesmente uma questão de substituições diretas de sequências de símbolos.

Diversos métodos e paradigmas têm sido propostos, desde mecanismos básicos, como simples substituição de palavras da língua fonte por palavras na língua alvo, até mais sofisticados como a criação manual de regras de tradução e geração automática de modelos estatísticos, conhecido como Tradução Automática Estatística (*Statistical Machine Translation - SMT*).

Atualmente, os sistemas de SMT baseada em frases⁴ são considerados o estado da arte na TA, de acordo com as métricas de avaliação automática BLEU (PAPINENI et al., 2002) e NIST (DODDINGTON, 2002). Ressalta-se que a SMT pode ser aplicada a praticamente qualquer par de línguas e tipo de *corpus* (conjunto de sentenças e suas traduções). O processo estatístico de tradução é baseado na construção do modelo de língua e modelo de tradução, que são construídos a partir do treinamento com *corpus* paralelos dos modelos da IBM (BROWN, 1993) que calculam a probabilidade de uma dada palavra ou frase fonte ser traduzida para uma palavra ou frase alvo. Segundo Caseli (2009), é possível ter resultados satisfatórios com *corpus* treinados com menos de um milhão de palavras em cada língua.

2.4 Ferramentas Computacionais de Apoio a Comunicação

Alguns exemplos de ferramentas que utilizam alguma forma de tradução automática são descritos a seguir.

Chen e Raman (2008) apresentam uma proposta que integra tradução automática em um chat por meio de um *bot*. Utilizando o Google Talk com integração ao Gmail, o usuário inicia uma sessão de *chat* com o *bot*, cujo endereço

⁴ Em SMT, frases são sequências de 2 ou mais palavras

se dá na forma “xx2yy@bot.talk.google.com” (onde xx é a sigla da língua fonte e yy é a sigla da língua alvo). Após enviar uma mensagem, o *bot* responde a essa entrada com a tradução da mesma. Tal interação pode ser observada na Figura 2-1.



Figura 2-1 *Chat com EN2ZH (CHEN e RAMAN, 2008)*

Polyglot (AIKEN e MINA, 2009) é um sistema que permite aos participantes escreverem comentários anonimamente e ser capaz de simultaneamente ler comentários de outros participantes de modo a suportar um grupo grande e multilíngue em uma conferência eletrônica. A interface do Polyglot para usuários cuja língua mãe é o inglês é apresentada na Figura 2-2. Utiliza o Google Translate para a TA em 41 idiomas.

Yamashita e Ishida (2006) exploram o impacto da TA no trabalho colaborativo entre as pessoas. Para realizarem uma tarefa colaborativa, os participantes utilizam o AnnoChat⁵ (Figura 2-3), que é capaz de prover informações sobre a escrita dos outros usuários, assim como, permitir que o usuário navegue pelos idiomas chinês, inglês, coreano e japonês. Por exemplo, um participante japonês que escolha o japonês como seu idioma para navegação e escrita, poderá ler e escrever em japonês. O AnnoChat utiliza um tradutor automático de uso comercial e possui recurso para anotações (similar a anotação de vídeos).

⁵ <http://yoshino.sys.wakayama-u.ac.jp/spark/?lng=en&page=AnnoChat>

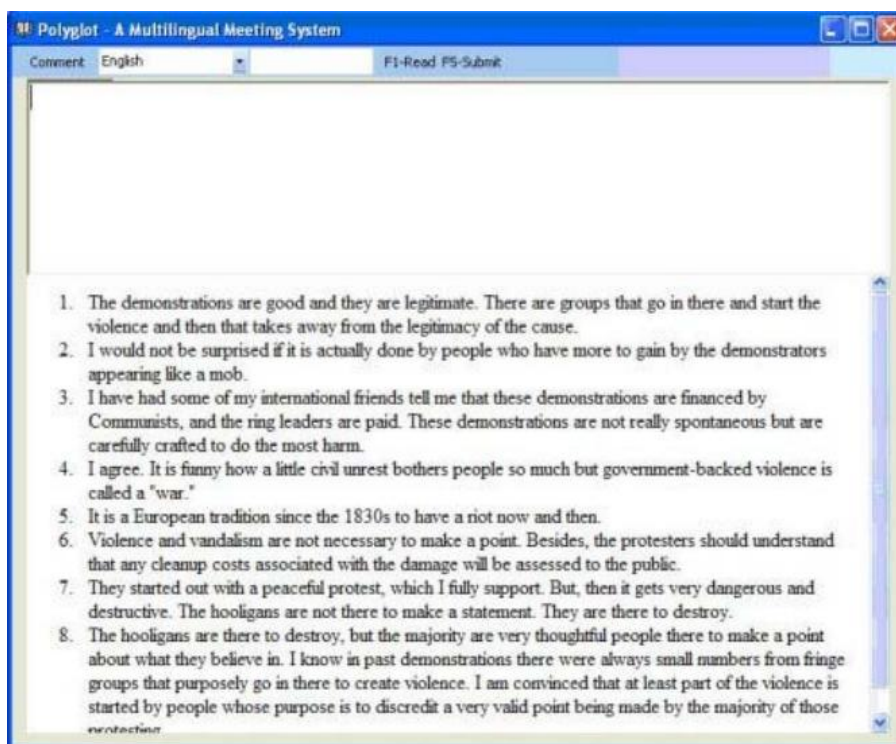


Figura 2-2 Interface do Polyglot para idioma inglês (AIKEN e MINA, 2009)



(a) Japanese Chat Interface

(b) Chinese Chat Interface

Figura 2-3 Interface do AnnoChat (YAMASHITA e ISHIDA, 2006)

O LiveMocha⁶ é uma rede social internacional para pessoas que querem aprender outras línguas. Ele contém lições audiovisuais e conta com a colaboração

⁶ <http://www.livemocha.com>

de outros usuários em um esquema de colaboração e aprendizado mútuo. Uma das ferramentas disponíveis é um *chat* que possui integrado com o Google Translate, onde o usuário pode realizar buscas por traduções de palavras.

O TBot⁷ é um *bot* para Windows Live Messenger que possui o mesmo mecanismo do trabalho de Chen e Raman (2009). O TBot utiliza o Bing Translator da Microsoft.

Uma comparação das características dos sistemas é apresentada na Tabela 2-1. Percebe-se que a TA tem sido usada de diversas maneiras para auxiliar a comunicação entre usuários, entretanto outros recursos, como por exemplo, aqueles que consideram a cultura dos usuários, não têm sido agregados às ferramentas de modo a facilitar essa comunicação.

Tabela 2-1 Comparação entre ferramentas com TA

	É pesquisa científica ?	Está disponível ?	É ferramenta Web?	Permite a comunicação entre usuários?	Permite a edição da TA?	Considera a cultura dos participantes ?
GTalk	Sim	Sim	Sim	Não diretamente	Não	Não
Polyglot	Sim	Não	Não	Sim	Não	Não
AnnoChat	Sim	Sim	Não	Sim	Não	Não
LiveMocha	Não	Sim	Sim	Sim	Não	Não
TBot	Não	Sim	Não	Não diretamente	Parcialmente	Não

Este trabalho propõe uma ferramenta que, além de apoiar a comunicação entre pessoas de diferentes culturas, utiliza um tipo de conhecimento cultural em conjunto com a TA para ampliar a ajuda ao usuário nessa comunicação. A próxima seção aborda um tipo de conhecimento cultural adotado neste trabalho.

⁷ <http://www.microsofttranslator.com/user/bot/>

2.5 Senso Comum

O senso comum é um tipo de conhecimento cultural que abrange o conjunto de fatos e crenças que aprendemos por meio de nossas observações, ensinamentos de nossos pais e familiares, do que a sociedade considera como aceitável ou não em um determinado lugar e período de tempo. Minsky (1988) define o senso comum como “as habilidades mentais que a maioria das pessoas compartilha”. Assim o termo senso comum é utilizado em referência ao conhecimento sobre os aspectos espaciais, físicos, sociais, temporais e psicológicos adquiridos ao longo da vida das pessoas.

Neste trabalho, é adotada a definição de senso comum do projeto OMCS-Br: um conjunto de fatos conhecidos pela maioria das pessoas, “abrangendo uma ampla parte das experiências humanas, conhecimento sobre aspectos espaciais, físicos, sociais, temporais e psicológicos do dia-a-dia dos seres humanos” (LIU e SINGH., 2004). Um exemplo de conhecimento de senso comum do Brasil que se pode citar é quando uma pessoa, para sanar as dores no fígado, faz um chá de boldo que era usado pelos seus antepassados, sem conhecer o princípio ativo das folhas e seu efeito nas doenças do fígado. Para aquele grupo específico, na época em que o fato foi obtido, o fato era de senso comum. Outros exemplos de senso comum, segundo Singh (2001), são: “toda pessoa é mais jovem que sua mãe”; “papel queima se jogado no fogo”; “as pessoas normalmente dormem durante a noite”; “as pessoas normalmente vão a festas para conhecer outras pessoas”.

2.5.1 O Projeto OMCS-Br

O projeto OMCS no Brasil (OMCS-Br) teve início a partir de uma parceria do LIA (Laboratório de Interação Avançada) com o Media Lab do MIT, com objetivo de registrar os fatos de senso comum de contribuintes brasileiros. Assim como o OMCS, o projeto *Open Mind Common Sense* no Brasil (OMCS-Br) leva em consideração o fato de que qualquer pessoa possui o senso comum que se deseja fornecer às máquinas (SINGH, 2001; LIU e SINGH, 2004), tornando a construção da base de conhecimento um trabalho colaborativo, envolvendo as pessoas ao redor do mundo nesse desafio, usando os recursos providos pela Web/Internet e adoção da

representação do conhecimento em língua natural. Desse modo, aproveitando os recursos da internet e os avanços das pesquisas na área de processamento de língua natural, foi lançado com a disponibilização de um site o projeto OMCS-Br⁸

A arquitetura do projeto OMCS-Br pode ser dividida em três partes (TSUTSUMI, 2006), sendo:

1. O site: A inserção de informações pelos colaboradores é feita através de um site na Internet, o qual utiliza a tecnologia Java Server Pages (JSP) e as informações coletadas são armazenadas em um banco de dados relacional implementado em MySQL;
2. A ConceptNet: É uma rede semântica gerada a partir das informações armazenadas no banco de dados, as quais passam previamente por um *parser* para normalização do texto;
3. A API: são funções utilizadas para a manipulação da ConceptNet.

Para compreender melhor essa arquitetura, podemos observar a Figura 2-4, com as três partes citadas que são detalhadas nas seguintes seções.

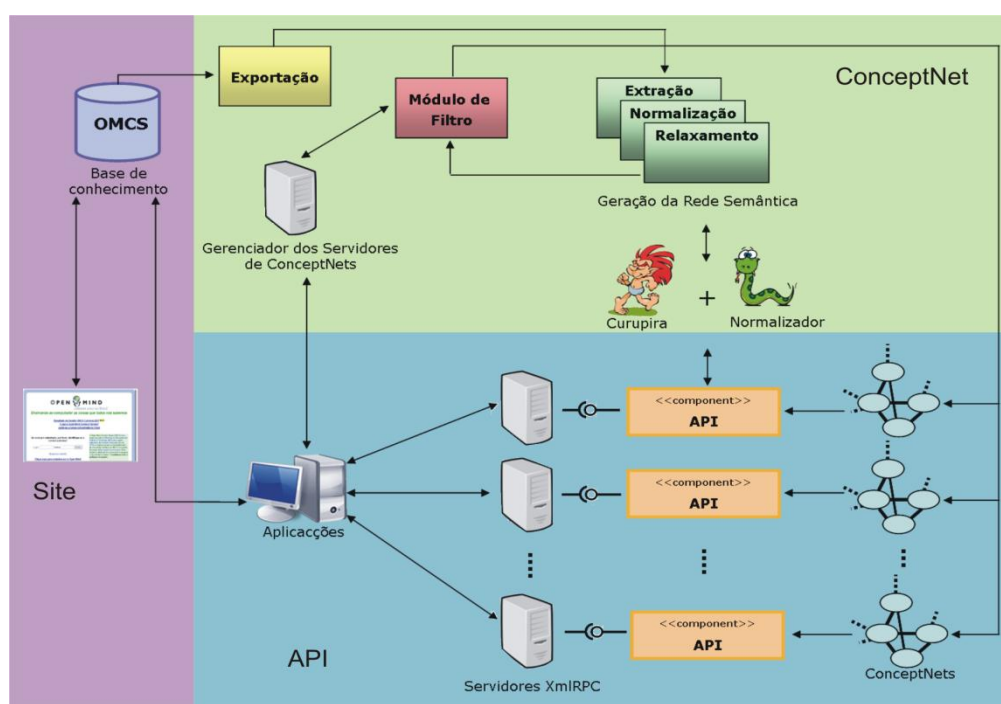


Figura 2-4 Arquitetura do projeto OMCS-Br

⁸ <http://www.sensocomum.ufscar.br>

2.5.1.1 O site

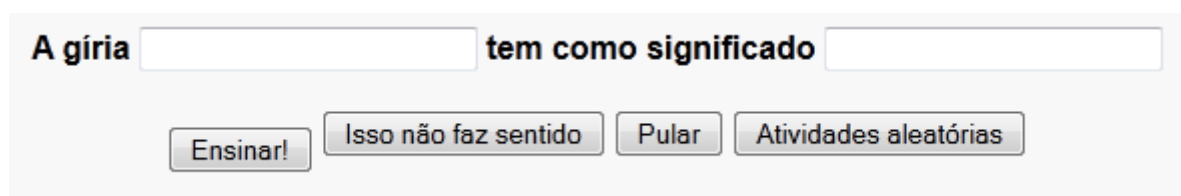
A melhor maneira encontrada para que voluntários de qualquer lugar do mundo possam contribuir com a inserção de fatos na base de conhecimento de senso comum foi a criação de um site. A contribuição é por meio de *templates*, que são frases com estruturas gramaticais simples, onde os colaboradores devem preencher lacunas. Os *templates* têm partes estáticas e dinâmicas. As partes dinâmicas mudam cada vez que são apresentadas e é resultado de uma retroalimentação de fatos já cadastrados na base.

Antes deste trabalho, a base de conhecimento do OMCS-Br não continha fatos relacionados à gírias, metáforas, expressões idiomáticas e outros termos cujo significado pode variar e depende da cultura dos usuários, denominados, neste trabalho, como expressões culturais. Para popular a base com tais informações, foi realizado o chamado Desafio de Carnaval, que teve duração de um mês, onde usuários de todo o Brasil foram convidados para inserir fatos relacionados às expressões culturais do país. Desafios como este são realizados de tempos em tempos para acelerar a coleta de um tema de interesse de pesquisa. Atualmente, estão cadastrado aproximadamente 2000 contribuidores do projeto OMCS-Br.

Desenvolvidos pelo pesquisador, alguns exemplos de *templates* são apresentados na Figura 2-5 e Figura 2-6. Outros *templates* criados são mostrados a seguir, sendo que a palavra em negrito representa a parte dinâmica:

- Uma gíria para **xxxx** é _____
- Um exemplo de gíria é o termo **xxxx**, que significa _____

Por meio destes *templates* foram coletadas aproximadamente 10.600 sentenças (sendo o total da base de dados do OMCS-Br de aproximadamente 300.000), que geraram por volta de 3300 expressões culturais diferentes. Exemplos de expressões culturais coletadas são: saracotear, desencanar, pegar leve, vazar, linguarudam, bater fio, dar o fora, lesma, cutucar onça com vara curta.



A gíria tem como significado

Figura 2-5 *Template* para coleta de gírias

A gíria **magrela** tem como significado

Figura 2-6 *Template* para coleta de significado de gírias

2.5.1.2 Módulo Gerador da ConceptNet

Cada um dos *templates* é relacionado a uma ou mais relações de Minsky, que são relações semânticas que conectam termos formando a ConceptNet. A ConceptNet é uma rede semântica do projeto OMCS-Br que serve para representar o conhecimento de maneira completa e eficiente, permitindo que novas informações sejam inferidas a partir de conhecimentos explícitos (LENAT et al, 1990). A ConceptNet é um conjunto de conceitos conectados através de um conjunto de relações definidas por Minsky para representar o conhecimento humano (MINSKY, 1988). Está formada por um conjunto de vinte relações, por exemplo: (isA "magrela" "gíria" "f=2;i=0"), (definedAs "magrela" "bicicleta" "f=1;i=0"), sendo f a quantidade de vezes que a relação foi gerada a partir de uma regra de extração e i a quantidade de vezes que a relação foi gerada a partir de uma regra de inferência.

Tabela 2-2 Os vinte tipos de relações usadas para construir a ConceptNet e seus respectivos exemplos (Liu et al., 2004)

Classe	Relação	Exemplo
K-lines	ConceptuallyRelatedTo	(ConceptuallyRelatedTo 'mau hálito' 'menta' 'f=4;i=0')
	ThematicKLine	(ThematicKLine 'vestido de noiva' 'véu' 'f=9;i=0')
	SuperThematicKLine	(SuperThematicKLine 'civilização ocidental' 'civilização' 'f=0;i=12')
Things	IsA	(IsA 'cavalo' 'mamífero' 'f=17; i=3')
	PropertyOf	(PropertyOf 'fogo' 'perigoso' 'f=17;i=1')
	PartOf	(PartOf 'borboleta' 'asa' 'f=5;i=1')
	MadeOf	(MadeOf 'bacon' 'porco' 'f=3;i=0')

	DefinedAs	(DefinedAs 'bife' 'carne de animal' 'f=2;i=1')
Agents	CapableOf	(CapableOf 'dentista' 'tirar dente' 'f=4;i=0')
Events	PrerequisiteEventOf	(PrerequisiteEventOf 'ler carta' 'abrir envelope' 'f=2;i=0')
	FirstSubeventOf	(FirstSubeventOf 'acender fogo' 'riscar isqueiro' 'f=2;i=3')
	SubEventOf	(SubEventOf 'praticar sport' 'pontuar' 'f=2;i=0')
	LastSubeventOf	(LastSubeventOf 'assistir ópera' 'aplaudir' 'f=2;i=1')
Spatial	LocationOf	(LocationOf 'exército' 'guerra' 'f=3; i=0')
Causal	EffectOf	(EffectOf 'ver video' 'entreterimento' 'f=2;i=0')
	DesirousEffectOf	(DesirousEffectOf 'suar' 'tomar banho' 'f=3;i=1')
Functional	UsedFor	(UsedFor 'lareira' 'queimar' 'f=1;i=2')
	CapableOfReceivingAction	(CapableOfReceivingAcion 'bebida' 'servir' 'f=0; i=14')
Affective	MotivationOf	(MotivationOf 'jogar' 'competir' 'f=3;i=0')
	DesireOf	(DesireOf 'pessoa' 'ser feliz' 'f=2;i=0')

Neste trabalho, há o interesse na relação “isA”, que representa uma relação de hierarquia entre os termos, e na relação “definedAs”, que conecta termos que possuem o mesmo significado.

2.5.1.3 A API

A Application Programming Interface (API) do projeto OMCS é uma interface que disponibiliza um conjunto de funções a outros softwares para que possam manipular a ConceptNet. A API pode ser dividida em dois módulos. Em um deles, composto por quatro funções: navegar, contexto, projeção e analogia as quais permitem buscas considerando uma entrada simples como “carro”, sendo que devem estar na forma canônica, ou seja, verbos no infinitivo e substantivos no singular. A Figura 3 apresenta, em uma interface visual disponibilizada pelo projeto, o resultado da função navegar para a entrada “carro”. Utilizando as funções descritas acima é possível identificar o contexto associado a algum conceito, realizar analogias e também fazer projeções sobre o conceito fornecido

No outro módulo composto pelas funções inferir conceito, inferir tópico, inferir humor e sumarizar, é feita uma busca utilizando como entrada sentenças inteiras, as quais são submetidas ao *parser*, que as divide e normaliza em estruturas menores para poderem ser tratadas como nodos na ConceptNet. Com essas funções é possível identificar o tópico principal do texto, classificá-lo em gênero, considerar o contexto para adquirir o sentido de uma palavra, fazer analogias para reconhecer novos conceitos e identificar o humor expresso no texto.

2.6 Considerações Finais

Este capítulo descreveu o campo de comunicação mediada por computador, como também, algumas ferramentas que possibilitam a comunicação entre pessoas de diferentes culturas. Visto que este tipo de comunicação pode ser difícil para alguns participantes, existem alguns recursos como a tradução automática para facilitar a tarefa deste usuário. Este trabalho propõe uma ferramenta que, além do uso tradução automática, adota um recurso que utiliza o senso comum de modo a torná-la mais contextualizada na cultura de seus usuários. Tal ferramenta será apresentada no capítulo seguinte.

Capítulo 3

CULTURE-TO-CHAT (C2C)

3.1 Considerações Iniciais

O capítulo a seguir descreve todas as funcionalidades do Culture-to-Chat (C2C) e, a forma com que ele disponibiliza as informações de senso comum. Além disso, este capítulo também relata o processo realizado para o desenvolvimento do protótipo e, as tecnologias utilizadas para sua implementação. Este capítulo está dividido em algumas seções que relatam: o modelo do processo de desenvolvimento do C2C (Seção 3.2); a arquitetura do C2C (Seção 3.3); o fluxo da informação no *chat* (Seção 3.4); a retroalimentação da base de conhecimento do OMCS-Br (Seção 3.5) e; as considerações finais (Seção 3.6).

3.2 Modelo de Desenvolvimento Baseado em Prototipação

Para o desenvolvimento de um *chat* que conta com apoio de uma base de conhecimento cultural, adotou-se um modelo de desenvolvimento baseado na abordagem proposta por Anacleto, Fels e Villena (2010). Esta abordagem é composta por ciclos ou etapas sendo que cada ciclo é composto por: objetivos, questões, protótipo, respostas e *stakeholders*. Por objetivo, entende-se como a meta ou tarefa a ser executada naquela etapa; as questões são perguntas que apareceram naquela etapa; o protótipo é composto por todos os recursos que

ajudaram na resolução das questões; as respostas são as soluções para as questões da etapa e os *stakeholder* são profissionais que auxiliaram o pesquisador naquela etapa.

Para o desenvolvimento do C2C, focou-se na prototipação como guia para cada etapa. Assim, o modelo de desenvolvimento do C2C é baseado em prototipação. Vários níveis de prototipação foram usados: baixa fidelidade, média fidelidade e alta fidelidade (ENGELBERG e SEFFA, 2002). Com isso, o modelo de desenvolvimento do C2C é composto por 5 etapas, cada uma contendo um protótipo. A Tabela 3-1 apresenta um resumo das 5 etapas que serão discutidas com mais detalhes adiante.

Tabela 3-1 Etapas do modelo de desenvolvimento do C2C

Etapa	Objetivo	Questão	Protótipo	Respostas	Stakeholders
1	Propor um mecanismo que combine a TA e o senso comum	Quais as funcionalidades da TA?	Primeiro protótipo em papel	Pesquisa bibliográfica da área	Docente de PLN
		Quais as funcionalidades do senso comum?		Pesquisa bibliográfica da área	
		Como juntar os dois recursos?		Proposta do protótipo	
2	Inserir o usuário no processo de tradução de mensagens	Qual modelo mental do usuário será mapeado?	Segundo protótipo em papel	Usuário remetente	Docente de PLN e 2 docentes de IHC
		Como será a interação do usuário?		Proposta do protótipo	
3	Aumentar o look-and-feel do protótipo	Quais ferramentas utilizar?	Protótipo de média fidelidade	Balsamiq Mockup e PowerPoint	Docentes de IHC, PLN e Redes
4	Utilizar os recursos existentes em um protótipo funcional	Como utilizar a API da ConceptNet?	Primeiro protótipo de alta fidelidade	Estudo dos trabalhos que utilizam senso comum	Docentes de PLN e IHC e pesquisadores de um congresso
		Como utilizar o Moses?		Relatórios da ferramenta	
5	Migrar o protótipo desktop para	Qual tecnologia utilizar?	Segundo protótipo de	Ruby on Rails dentre outras tecnologias web	Pesquisadores do LIA, docentes

	web	Como está a usabilidade do protótipo?	alta fidelidade	Testes com usuários	de IHC e PLN
		Como avaliar a ferramenta?		Estudo piloto	

O objetivo da primeira etapa era elaborar uma proposta que combinasse a tradução automática com o conhecimento cultural, instanciado aqui no senso comum. Algumas questões que foram levantadas nesta etapa estavam relacionadas à maneira com que o senso comum iria atuar em conjunto com a tradução automática, a ferramenta que iria utilizar esses recursos e como seria a interação do usuário com essa ferramenta. A partir dessas questões, foi desenvolvido o primeiro protótipo do trabalho, apresentado na Figura 3-1. Utilizou-se a prototipação em papel (SNYDER, 2003), considerada de baixa fidelidade, por esta ter características como rapidez e facilidade de mudança em etapas iniciais de projetos.

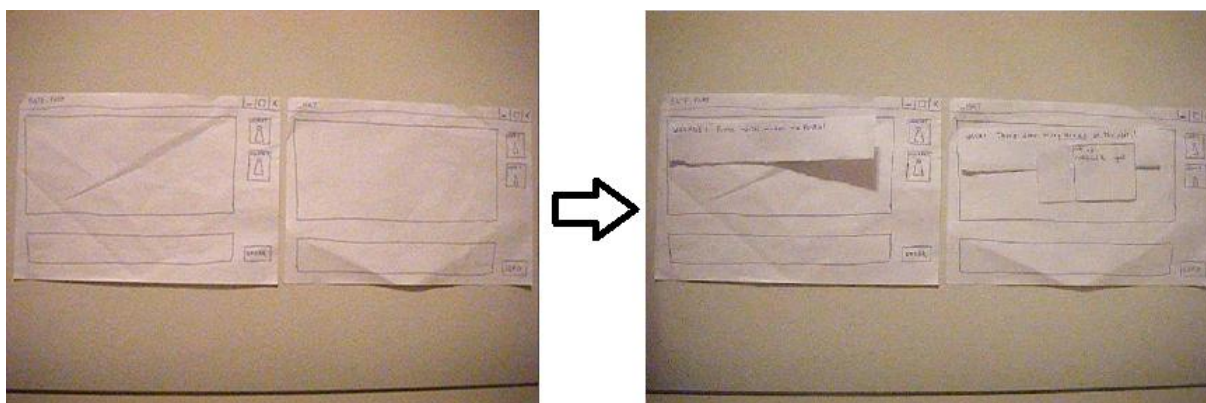


Figura 3-1 Primeiro protótipo do chat

Como primeira proposta, este protótipo ilustra o envio e recebimento de mensagens com algumas peculiaridades: o usuário remetente envia uma mensagem em sua língua, que é traduzida e recebida pelo usuário destinatário. Nessa simulação considerou-se que o destinatário contaria com auxílio de senso comum para palavras não traduzidas. Este protótipo, portanto, foca no modelo mental (COOPER, REIMANN e CRONIN, 2007) do remetente. O *stakeholder* desta etapa foi um docente da área de PLN que é especialista na área de TA, oferecendo maior apoio teórico sobre a área.

A segunda etapa teve início a partir da ajuda de um docente da área de IHC e do docente da área de PLN (*stakeholders*). Nessa segunda etapa, decidiu-se alterar

o foco para o modelo mental do usuário remetente. A Figura 3-2 apresenta o segundo protótipo do *chat*, também de baixa fidelidade, que recebeu o nome de “2-Chat”. Nele o usuário remetente conta com recursos de tradução automática e sugestões de senso comum.

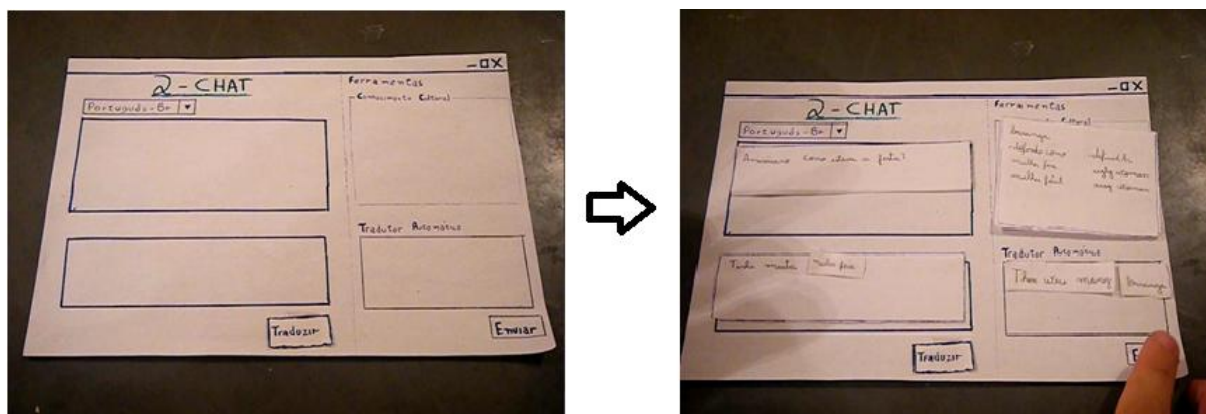


Figura 3-2 Segunda versão do protótipo, agora denominado 2-chat

O terceiro protótipo e, portanto, a terceira etapa teve como objetivo evoluir o protótipo da etapa anterior, desenvolvendo um protótipo de média fidelidade, a fim de aumentar o *look-and-feel* da interface e ainda manter as características dos protótipos de baixa fidelidade. Nota-se que nesta nova versão do protótipo o nome da aplicação foi alterado para Culture-to-Chat (C2C), mantido até o final do desenvolvimento. Na construção desse protótipo foi utilizado o Balsamiq Mockup⁹ em conjunto com Microsoft Office PowerPoint¹⁰. Tal protótipo é mostrado na Figura 3-3. Os *stakeholders* desta etapa compõe uma banca de docentes de diversas áreas da computação (IHC, PLN, Redes) que validaram a proposta.

O protótipo descrito apresentava a ideia básica do C2C, porém não era possível utilizá-lo de maneira funcional. Para possibilitar a simulação das funcionalidades do *chat* foi construído o protótipo mostrado na Figura 3-4, utilizando a linguagem Java (J2SE). Com este protótipo, é possível fazer busca por palavras relacionadas na base de conhecimento do projeto OMCS-Br por meio da ConceptNet. O tradutor automático utilizado foi o SMT Moses Toolkit¹¹, sendo que o *corpus* disponível para treinamento deste tradutor continha artigos da versão *online*

⁹ <http://www.balsamiq.com/builds/mockups-web-demo/>

¹⁰ <http://office.microsoft.com/pt-br/downloads/CD010200683.aspx>

¹¹ <http://www.statmt.org/moses/>

da revista científica Pesquisa FAPESP¹² (CASELI, SUGIYAMA e ANACLETO, 2010).

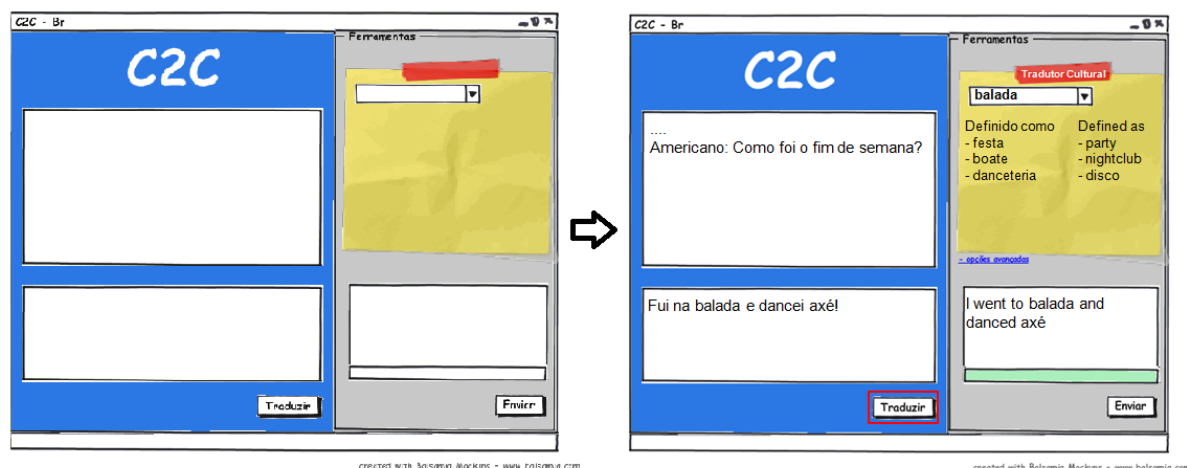


Figura 3-3 Protótipo de média fidelidade para o chat, agora denominado C2C

O protótipo apresentado foi construído com o propósito de demonstração das funcionalidades de senso comum e tradução automática, mas não realizava, ainda, a comunicação entre pessoas (remetente e destinatário). O protótipo de média fidelidade é para dar aos usuários, neste caso aos stakeholders, o “look and feel” que a ferramenta implementada terá. As tecnologias envolvidas nesta etapa foram escolhidas devido à facilidade e rapidez de desenvolvimento. Os principais *stakeholders* desta etapa foram o docente de PLN já mencionado anteriormente, assim como, outros dois docentes da área de IHC.

Com o objetivo de portar o protótipo *desktop* para uma versão *web*, inicia-se a quinta etapa do desenvolvimento do C2C. Inicialmente, algumas questões levantadas nesta etapa estavam relacionadas às tecnologias envolvidas nesta migração. O framework Ruby on Rails¹³ foi escolhido como plataforma de desenvolvimento da versão web do C2C, devido ao alto desempenho de aplicações que usam tal tecnologia e também para exploração da mesma e aprendizado.

¹² <http://revistapesquisa.fapesp.br>

¹³ <http://rubyonrails.org/>

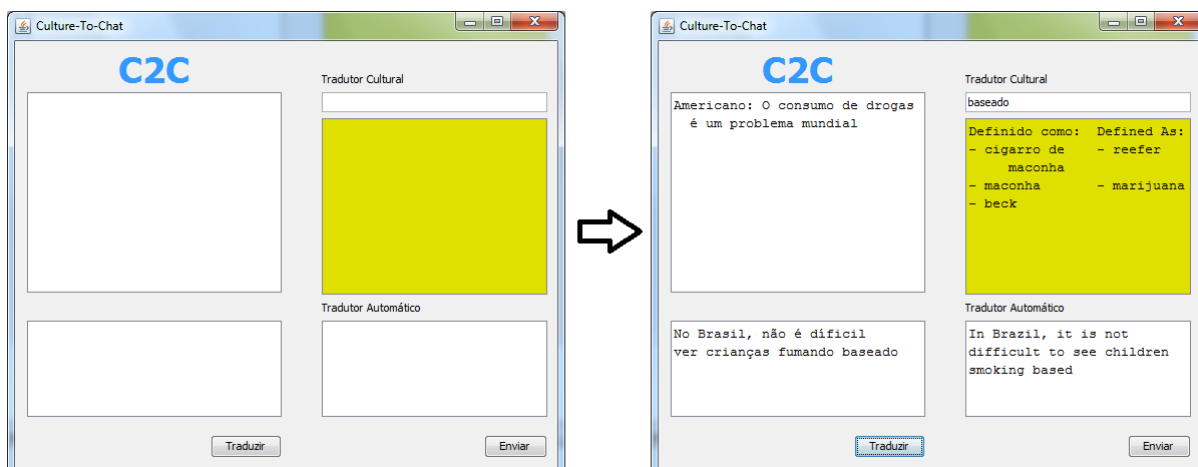


Figura 3-4 Primeiro protótipo de alta-fidelidade do C2C

A Figura 3-5 apresenta as telas iniciais do C2C¹⁴ para que o usuário realize o *login* e entre na sala de bate-papo. Nota-se que a base de dados de usuários do projeto OMCS-Br está interconectada com a do C2C, ou seja, usuários cadastrados em uma base também estão cadastradas na outra.

Na sala de bate-papo, é exibida uma lista dos usuários online. Ao clicar no nome de algum usuário é exibida a tela apresentada na Figura 3-6.

Na Figura 3-6, o item I corresponde à área da conversa, ou seja, traz as mensagens trocadas entre remetente e destinatário. O item II representa a caixa de texto para a escrita de mensagens em português (língua do remetente). Ao se clicar no botão “Traduzir” (item III), o tradutor cultural (item IV) pode apresentar algumas sugestões relacionadas às palavras da mensagem em português e o tradutor automático (item V) apresenta uma possível tradução gerada automaticamente para a sentença presente na área II. O botão “Enviar” (item VI) manda a mensagem contida na caixa de texto do tradutor automático para o destinatário.

¹⁴ <http://lia.dc.ufscar.br:3000/c2c>

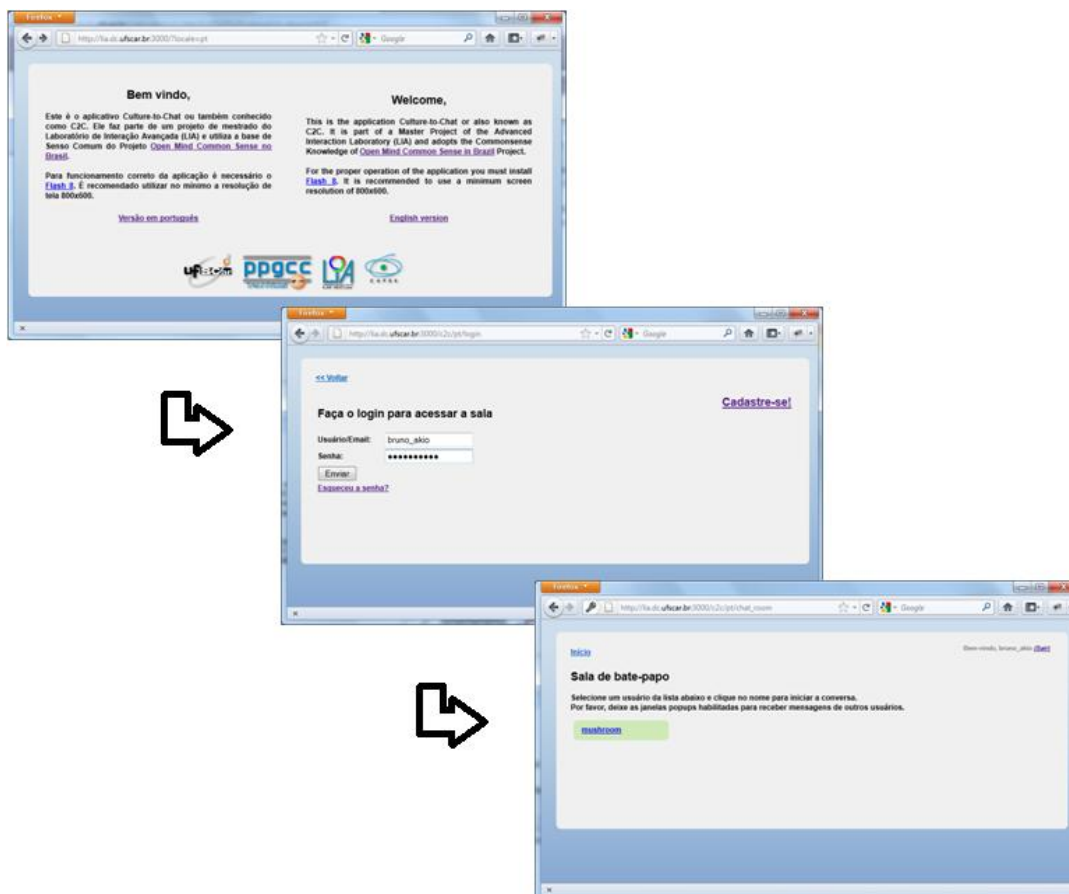


Figura 3-5 Entrando na sala de bate-papo do C2C

Com relação aos protótipos anteriores, a versão web apresenta importantes diferenças relacionadas ao tradutor cultural e ao tradutor automático. Quando o usuário pede para traduzir a mensagem em português, o tradutor cultural executa o seguinte algoritmo para identificação de palavras que podem ser uma expressão cultural (como gírias, metáforas ou expressão coloquial): Para cada palavra da mensagem escrita no item II da Figura 3-6, utiliza-se a ConceptNet para identificar quais estão conectadas por meio de uma relação “*isA*” com o termo “gíria” (que engloba não somente gírias, mas também outras expressões culturais); após a identificação das expressões culturais, é feita uma nova busca na ConceptNet, procurando retornar os termos conectados pela relação “*definedAs*”. Estes termos são considerados sinônimos culturais. A busca direta por termos continua disponível, ou seja, além do algoritmo de busca automática para expressões culturais, o próprio usuário pode inserir a palavra que deseja consultar na base de senso comum no item IV da Figura 3-6 e clicar no botão “Pesquisar”.

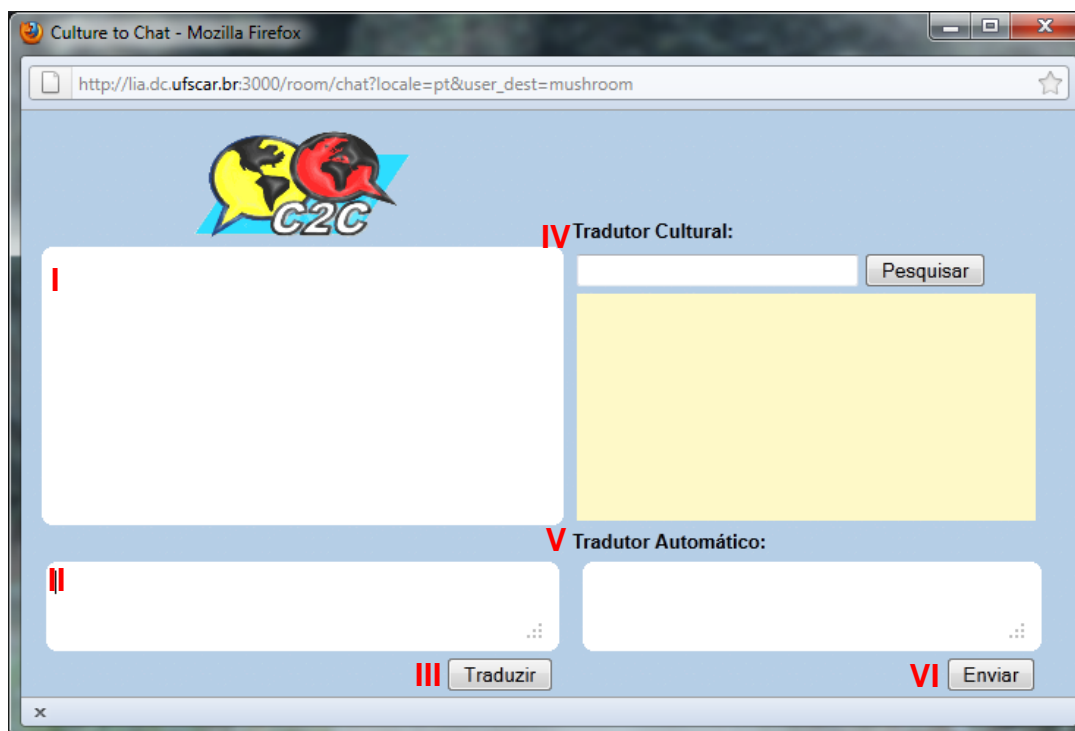


Figura 3-6 Tela inicial do C2C

Neste protótipo, optou-se pela troca do SMT Moses Toolkit para o Google Translate¹⁵, visto que o tamanho do *corpus* disponível para o primeiro tradutor, aproximadamente 500.000 palavras em cada idioma (português e inglês), segundo Caseli e Nunes (2009), é consideravelmente menor que o tamanho do corpus do segundo tradutor. Outra mudança ocorre com relação à comunicação com um usuário estrangeiro (destinatário). Decidiu-se que a conversa seria realizada em inglês (diferentemente da ideia inicial, onde cada usuário receberia mensagens em sua língua nativa) e, portanto, para usuários estrangeiros, a interface do C2C é similar a de outros chats (Figura 3-7).

Um exemplo de interação de um usuário brasileiro com o C2C é apresentado na Figura 3-8. Durante uma conversa, o usuário brasileiro fez uso da seguinte sentença: “Minha magrela foi roubada”. Ao mandar a mensagem ser traduzida, o tradutor cultural, utilizando o processo descrito anteriormente nesta seção, consegue identificar que o termo “magrela” pode ter um sentido coloquial, ou seja, existe na ConceptNet a relação “isA(magrela, gíria)” e sugere alguns sinônimos X, onde X está em “definedAs(magrela, X)”. O tradutor automático oferece a tradução “*My skinny was stolen*”, que neste caso não representa o mesmo sentido da sentença original

¹⁵ <http://translate.google.com.br/>

do brasileiro. A partir das sugestões de ambos os recursos (tradutor cultural e tradutor automático), o usuário brasileiro pode editar tanto a sentença original (e novamente traduzi-la) quanto a sentença traduzida antes de enviá-la ao destinatário.



Figura 3-7 Interface do C2C em inglês para usuários estrangeiros



Figura 3-8 Exemplo de interação do usuário brasileiro no C2C

Como exemplo de edição e posterior envio da mensagem em inglês, é apresentada a Figura 3-9. O usuário decidiu substituir a palavra “skinny” pela tradução da sugestão oferecida pelo tradutor cultural (“bike”) e, após essa modificação, enviou a mensagem para o outro usuário. Ressalta-se que outra possibilidade seria a edição da sentença em português e em seguida o início do processo de tradução e envio dessa nova mensagem.

Destaca-se que esta etapa foi a mais longa das etapas no ciclo de vida executado, conforme visto na Tabela 3-1, e que os *stakeholders* são compostos pelos docentes de IHC e PLN já citados como também os pesquisadores do LIA.

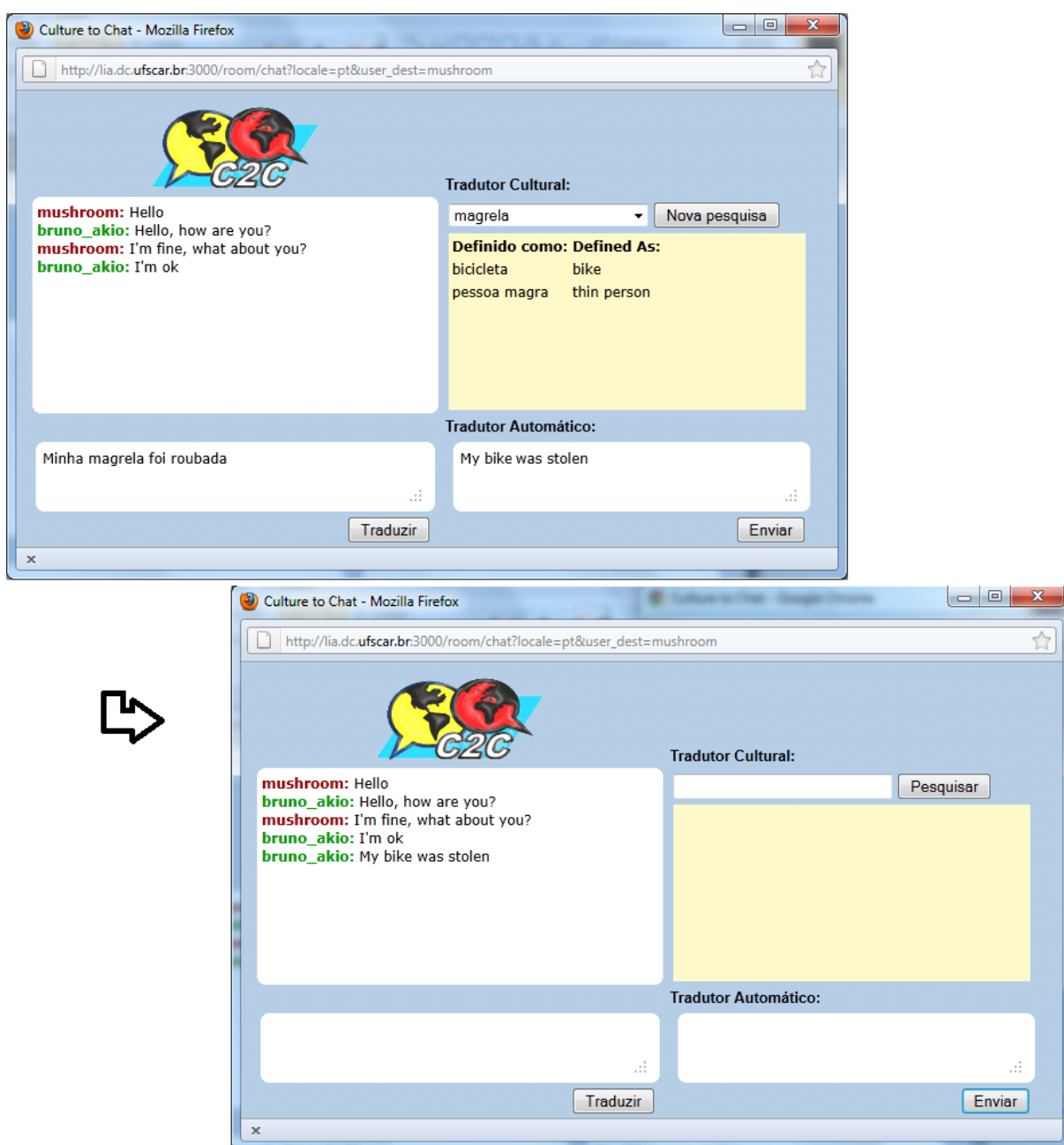


Figura 3-9 Edição e envio da mensagem em inglês no C2C

3.3 A Arquitetura do C2C

A Figura 3-10 apresenta a arquitetura do C2C (protótipo web). Os dados de entrada do usuário, via requisição HTTP (*Hypertext Transfer Protocol*) ou chamada AJAX (*Asynchronous Javascript and XML*), são enviados ao servidor de aplicação (Mongrel), que contém as definições dos modelos, visões e controles do C2C. O Mongrel tem acesso ao banco de dados do OMCS-Br e do C2C, ambos em MySQL. O servidor de aplicação faz chamadas à ConceptNet por meio de uma API implementada em Java e XML (*eXtensible Markup Language*). O Google Translate é acessado por meio de um *web service* disponível em um pacote (*gem*) do Ruby on Rails. O Juggernaut¹⁶ é responsável pelo envio/saída de dados do servidor para os clientes. Ele é considerado um *push server*, isto é, permite “empurrar”/mandar dados para os clientes por meio de uma aplicação web, provendo uma conexão em tempo real entre servidor e *browser* do usuário.

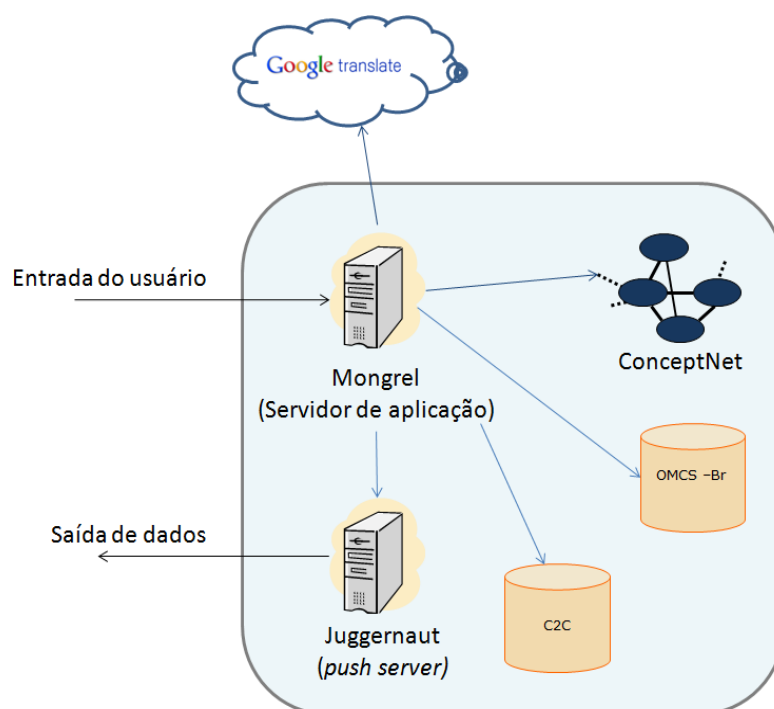


Figura 3-10 Arquitetura do C2C

¹⁶ <http://juggernaut.rubyforge.org/>

3.4 O Modelo de Fluxo da Informação do C2C

A partir da interação do usuário e da arquitetura do sistema é possível traçar o fluxo da informação do C2C no processo de criação de mensagens em língua estrangeira. Na Figura 3-11, O usuário brasileiro envia uma mensagem em português ao servidor (seta I), que manipula essa mensagem e executa chamadas à ConceptNet e ao Google Translate a fim de criar a lista de sugestões e traduções (setas II). Essa informação é enviada de volta para o usuário brasileiro (seta III) que pode editar a mensagem e retorná-la ao servidor (seta IV), para que este encaminhe ao destinatário correspondente (setas V). No caso dos usuários estrangeiros, a mensagem é enviada para o servidor que encaminha para o destinatário, nesse caso, sem nenhum processamento extra e, portanto não ilustrada na Figura 3-11.

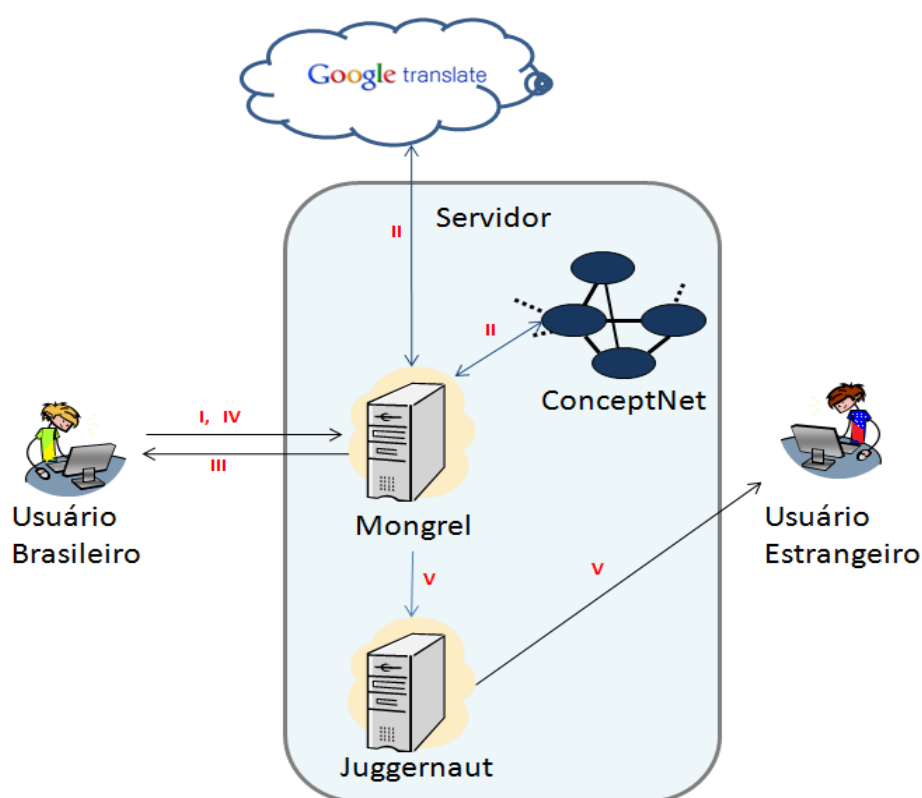


Figura 3-11 Fluxo da informação no envio da mensagem do usuário brasileiro para o usuário estrangeiro

3.5 O Processo de Retroalimentação da Base de Conhecimento do OMCS-Br

Para contribuir com o projeto OMCS-Br, o C2C conta com um mecanismo de retroalimentação que adiciona relações na base de conhecimento do OMCS-Br. Quando o usuário envia uma mensagem, é acionado, em segundo plano, um algoritmo que procura extrair relações da interação do usuário. A seguir, é descrito o processo de retroalimentação pelo C2C.

Neste processo, uma interação pode ser definida como a tradução de uma mensagem em português, a edição da mensagem traduzida e o envio desta última. Quando o usuário faz uma pesquisa direta no tradutor cultural, o termo pesquisado é guardado em uma variável de sessão. Os termos providos pelas sugestões automáticas do tradutor cultural também são armazenados nessa variável. Ao enviar a mensagem em inglês, é feita uma verificação se a sentença traduzida contém alguma das traduções dos sinônimos culturais sugeridos. Em caso afirmativo, é criada uma relação “*definedAs*” conectando o termo pesquisado e seu sinônimo cultural. Tal relação é inserida na base de conhecimento do projeto OMCS-Br.

No exemplo da Figura 3-8, ao traduzir a sentença “Minha magrela foi roubada”, o termo “magrela” é guardado em uma variável de sessão. Supondo que foi realizada a alteração da tradução da mensagem conforme mostrado na Figura 3-9, a fim de corrigir o sentido de sua mensagem, o processo de retroalimentação tem como parâmetros a variável de sessão e a mensagem final em inglês. Quando o processo é iniciado, a partir do envio da mensagem, é verificado se a sentença “*My bike was stolen*” contém as expressões “*bike*” ou “*thin person*”, respectivas traduções dos termos “bicicleta” e “pessoa magra” (sinônimos culturais da palavra “magrela”). Como a sentença contém a palavra “*bike*”, pode-se pressupor que o usuário concorda que “bicicleta” é um sinônimo cultural de “magrela”. É criada, então, uma sentença que irá gerar uma nova relação “*definedAs(magrela, bicicleta)*” a qual é inserida na base de conhecimento do projeto OMCS-Br.

Tal processo de retroalimentação, impacta tanto na base do OMCS-Br quanto no próprio C2C. Com a inserção dessa relação na base, a variável *f* (frequência) será acrescida. A ordem das sugestões no tradutor cultural é baseada no valor da variável de frequência (ordem decrescente). Portanto, as sugestões mais usadas

aparecerão no topo da lista do tradutor cultural e quanto mais o C2C é usado, mais próximo ao contexto do usuário estarão as sugestões.

3.6 Considerações Finais

O capítulo 3 abordou os recursos existentes no C2C para auxiliar o usuário na escrita de mensagens em língua estrangeira. De modo a apresentar as funcionalidades, como também, algumas possibilidades de utilizar o conhecimento de senso comum, foi apresentado um exemplo de escrita, tradução e edição de mensagem. Por meio dessa interação, observou-se que o C2C tem o potencial de auxiliar o usuário na comunicação entre usuários de diferentes culturas. De modo a investigar e observar o uso desta ferramenta, foi realizado um estudo piloto envolvendo usuários brasileiros e canadenses, descrito no próximo capítulo.

Capítulo 4

ESTUDO PILOTO

4.1 Considerações Iniciais

Este capítulo descreve o estudo piloto realizado com usuários canadenses e brasileiros com o intuito de avaliar o uso do C2C na comunicação entre pessoas de diferentes culturas e coletar a opinião desses usuários sobre o mesmo. Para realizar o estudo de caso foram definidas três etapas, cada uma delas se encontra em uma seção deste capítulo, como pode ser observado a seguir: Seção 4.2 descreve o local de estudo; Seção 4.3 relata sobre o planejamento do estudo piloto; Seção 4.4 apresenta os objetivos das etapas definidas para o estudo piloto, nesse caso duas etapas que variaram em termos da utilização (Etapa 1) ou não (Etapa 2) do conhecimento no C2C; Seção 4.5 explica os questionários utilizados para avaliar o C2C; Seção 4.6 aborda a primeira etapa; Seção 4.7 descreve a segunda etapa; Seção 4.8 faz um comparativo entre as duas etapas e; Seção 4.9 apresenta as considerações finais.

4.2 Local de Estudo

Devido a parceria do LIA/UFSCar com o laboratório MAGIC na Universidade de British Columbia no Canadá, foi possível a participação deste aluno no projeto

ELAP (*Emerging Leaders in the Americas Program*¹⁷), proposto pelo governo canadense, cujo propósito é apoiar o desenvolvimento de capital humano e a próxima geração de líderes nas Américas, estreitando os laços entre as universidades Canadenses e as instituições na America Latina e Caribe. O aluno fez um estágio de 6 meses no MAGIC Lab - *University of British Columbia* entre fevereiro e agosto de 2011, tendo se envolvido em várias pesquisas em andamento naquele laboratório de pesquisa. Nesse tempo, foi realizado o estudo piloto deste projeto, que envolveu usuários canadenses da cidade de Vancouver (British Columbia, Canadá) e usuários brasileiros de cidades de Santos, São Carlos e Taubaté do estado de São Paulo. Cada usuário participou do estudo de caso do local mais conveniente para o mesmo, no qual se inclui o local de trabalho (por exemplo, laboratório de pesquisa), local de estudo (universidade) e residência.

4.3 Planejamento

Objetivo do estudo piloto: investigar o uso do C2C na comunicação entre brasileiros e canadenses, como também, verificar se o senso comum apoia o usuário brasileiro na criação de mensagens em inglês.

Hipóteses:

1. O C2C oferece um canal de comunicação para usuários de diferentes localidades e idiomas
2. O senso comum apoia o usuário brasileiro na criação de mensagens em uma segunda língua
3. Os recursos do C2C ajudam o usuário na comunicação em inglês

Método: Para provar as hipóteses, realizam-se sessões de conversa entre usuários, sendo um de nacionalidade brasileira (nativo do português) e outro de nacionalidade canadense (nativo do inglês), por meio de uma ferramenta computacional. A comunicação é feita em inglês, sendo que para os usuários brasileiros, a ferramenta computacional oferece mecanismos de auxílio na criação

¹⁷ http://www.scholarships-bourses.gc.ca/scholarships-bourses/non_can/elap-pfla.aspx?lang=eng&view=d

da mensagem na língua estrangeira. Ambos os usuários têm acesso a um conjunto de questões como sugestão de conversa (Apêndice E). Para tanto:

1. Foi utilizado o protótipo do *chat* “C2C” desenvolvido pelo pesquisador do Laboratório de Interação Avançada do DC-UFSCar (LIA/DC-UFSCar), autor da dissertação;
2. Os mecanismos de auxílio da ferramenta computacional consistem em um tradutor automático e uma base de conhecimento cultural (descritos no Capítulo 3);
3. O conhecimento de senso comum foi disponibilizado ao usuário brasileiro para sugerir palavras ou expressões alternativas à mensagem do mesmo ou para oferecer uma ampliação do vocabulário por meio das traduções das sugestões.
4. As sugestões de conversa foram elaboradas pelo pesquisador baseadas em atividades de aulas de *English as a Second Language* (ESL), conhecidas como *warm-up activities* (HILL e FLYNN, 2006);
5. Os dados do estudo piloto foram coletados por meio de questionários, capturas de telas e observação da interação dos usuários. Para a captura de tela, era necessário que o usuário fizesse o *download* de um programa para transmissão remota.

Seleção e Perfil dos Participantes: O perfil dos prováveis usuários brasileiros do C2C engloba jovens (adolescentes ou adultos) com entendimento no mínimo básico da língua estrangeira (no caso, o inglês). Visto que os jovens, em geral, apresentam uma linguagem peculiar composta por gírias, coloquialismos e neologismos (PIMENTEL, 2006), acredita-se que esse perfil se encaixa com a proposta do C2C. Um entendimento básico de inglês também é necessário visto que a comunicação é feita nesta língua. Para o perfil dos usuários falantes da língua inglesa, escolhem-se jovens também, pois mesmo em diferentes culturas, os jovens tendem a ter preferências semelhantes (KJELDGAARD e ASKEGAARD, 2006).

Para este estudo piloto, voluntariaram-se 8 (oito) pares de participantes, cada par composto por um brasileiro e um canadense (16 usuários no total). Os usuários brasileiros pertenciam à faixa etária de 22 a 27 anos enquanto que os usuários canadenses pertenciam à faixa etária de 24 a 29 anos. O nível de escolaridade dos usuários variava entre superior incompleto, superior completo e pós-graduação.

4.4 Etapas do Estudo Piloto

Como já discutido, o estudo piloto é um experimento preliminar realizado para avaliar a viabilidade, o tempo, o custo, os eventos e efeitos antes de um estudo de maior escala.

Este estudo foi dividido em duas etapas, ambas compostas pelos seguintes procedimentos:

- Apresentação da ferramenta C2C com exemplos de suas funcionalidades;
- Envio do questionário pré-sessão para os participantes;
- Envio das questões de sugestão de conversa (*warm-up questions*);
- Configuração para gravação da interface do usuário brasileiro;
- Realização da sessão de bate-papo por meio do C2C;
- Envio do questionário pós-sessão para os participantes.

A diferença entre as duas etapas do estudo piloto está na disponibilidade dos recursos do C2C. Na primeira etapa, todos os recursos estão disponíveis ao usuário como descrito na seção 3.3. Na segunda etapa, a funcionalidade de tradução cultural foi omitida a fim de se fazer posteriormente uma análise comparativa entre as duas etapas. Tal interface criada para a Etapa 2 é apresentada na Figura 4-1.

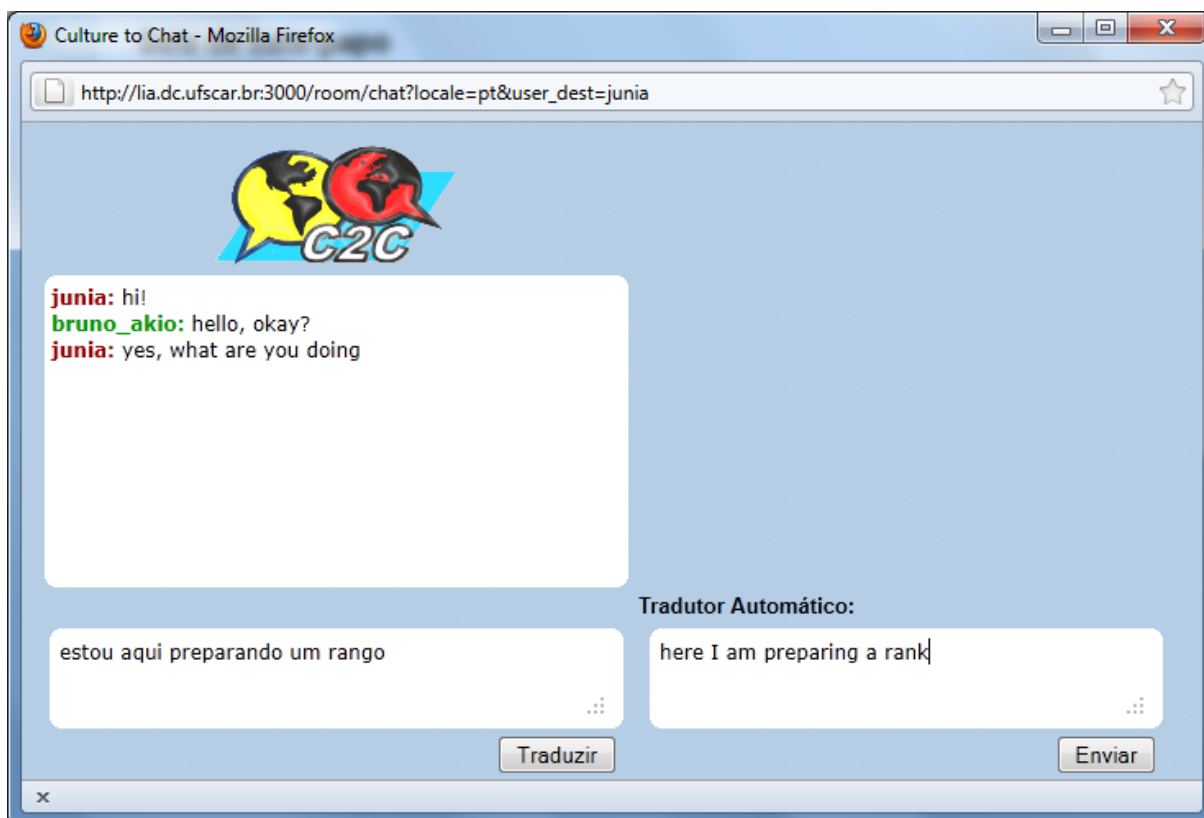


Figura 4-1 Interface do C2C sem o recurso de tradução cultural utilizada na Etapa 2 do estudo piloto

4.5 Preparação dos Questionários Utilizados no Estudo Piloto

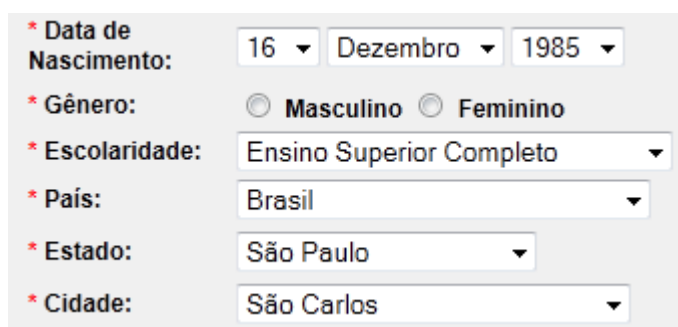
Nesta seção, os questionários elaborados para os usuários brasileiros são apresentados e, em seguida, os questionários aplicados com os usuários canadenses são descritos. Ressalta-se que as figuras a seguir ilustram apenas parte dos questionários. Contudo, nos apêndices informados durante o texto há os questionários completos.

4.5.1 Questionário Pré-Sessão: Usuário Brasileiro

Este questionário (Apêndice A), aplicado no início de cada etapa do estudo piloto, tem como objetivo coletar o perfil do usuário brasileiro. Nota-se que se pode obter outras informações referentes ao perfil dos usuários a partir do cadastro dos mesmos na base de dados do projeto OMCS-Br (requisito necessário para a entrada na sala de bate-papo do C2C). Portanto, este questionário e o cadastro do usuário

são informações complementares do perfil. O perfil do usuário é importante para o processo de retroalimentação descrito na Seção 3.5, pois a extração de conhecimento de senso comum está relacionada a ele.

Sendo assim, algumas informações importantes coletadas no cadastro são a idade, a escolaridade e a região de moradia, como pode ser observado na Figura 4-2.



* Data de Nascimento: 16 ▾ Dezembro ▾ 1985 ▾

* Gênero: Masculino Feminino

* Escolaridade: Ensino Superior Completo ▾

* País: Brasil ▾

* Estado: São Paulo ▾

* Cidade: São Carlos ▾

Figura 4-2 Exemplo de alguns campos do cadastro do usuário

Por meio do questionário, pretende-se também identificar a familiaridade do usuário brasileiro com: o idioma estrangeiro, o uso do computador para se comunicar, o uso do computador para se comunicar em inglês. Algumas perguntas com esse objetivo podem ser observadas na Figura 4-3.

Estas questões, como as apresentadas na Figura 4-3, foram feitas porque podem influenciar diretamente nos resultados do estudo piloto. O grau de dificuldade no uso da ferramenta ou a interação podem ser influenciados se o usuário não utiliza o computador para comunicação. Com relação ao nível de inglês, tais dados servem de comparação com o questionário pós-sessão do usuário canadense, a ser apresentado na seção 4.5.4.

Você estuda ou já estudou Inglês?

Sim

Não

Como você considera o seu nível de Inglês?

Avançado

Intermediário

Básico

Other:

Você usa o computador para se comunicar (msn, salas de bate-papo etc)?

Muito frequentemente (diariamente)

Frequentemente (semanalmente)

Indiferente

Poucas vezes (mensalmente)

Raramente (semestralmente)

Não tenho como opinar

Com relação a escrita e leitura em inglês, que nota você daria para si?

selecione uma opção ▼

selecione uma opção

10

9

8

7

6

5

4

3

2

1

0

Já utilizou o computador para aprender/estudar Inglês?

Sim

Não

Figura 4-3 Exemplos de perguntas do questionário pré-sessão do usuário brasileiro

4.5.2 Questionário Pós-Sessão: Usuário Brasileiro

Ao final da realização da sessão de bate-papo, foi aplicado o questionário pós-sessão do usuário brasileiro (Apêndice B) de modo a coletar informações referentes à sessão de bate-papo do ponto de vista do usuário. Algumas perguntas são apresentadas na Figura 4-4 Exemplos de perguntas do questionário pós-sessão do usuário brasileiro.

Como foi a interação com o C2C

Muito fácil

Fácil

Indiferente

Difícil

Muito difícil

Não posso opinar

O que você achou do desempenho/velocidade do C2C?

Muito rápido

Rápido

Indiferente

Lento

Muito lento

Não posso opinar

O que você achou do tradutor automático no chat?

Muito útil

Útil

Indiferente

Inútil

Muito inútil

Não posso opinar

O que você achou das sugestões do tradutor cultural?

Muito útil

Útil

Indiferente

Inútil

Muito inútil

Não posso opinar

Figura 4-4 Exemplos de perguntas do questionário pós-sessão do usuário brasileiro

O intuito destas perguntas é oferecer ao usuário brasileiro a oportunidade de relatar a experiência com o C2C nos quesitos de usabilidade, desempenho e funcionalidade, como também, descrever a experiência de comunicação com um usuário estrangeiro e sugerir melhorias para a ferramenta.

Na segunda etapa do estudo piloto, houve uma alteração na pergunta relacionada às sugestões do tradutor cultural. Como essa funcionalidade foi omitida, os usuários não têm como opinar sobre tal. No caso, a pergunta foi substituída por: “O que você acha se o C2C tivesse mecanismos para ajudar na tradução de expressões culturais (gírias, coloquialismos etc.)?”. Nesse caso, tal pergunta ajuda a identificar o quanto os usuários sentem necessidade de recursos adicionais nas ferramentas mediadoras de comunicação.

4.5.3 Questionário Pré-Sessão: Usuário Canadense

Este questionário (Apêndice C), que foi aplicado nas duas etapas, tem como objetivo investigar: a experiência dos usuários com o uso de ferramentas computacionais de comunicação e com a comunicação com pessoas de outras culturas, como por exemplo, os brasileiros. Algumas perguntas estão na Figura 4-5. Assim como no questionário pré-sessão do brasileiro este teve como intuito coletar informações dos usuários sobre experiências e conhecimentos que pudessem influenciar no resultado do estudo piloto.

Do you use a computer to interact with other people?	Are you interested to interact with foreigners?
<input type="radio"/> Always (daily)	<input type="radio"/> Yes
<input type="radio"/> Usually (weekly)	<input type="radio"/> No
<input type="radio"/> Indifferent	
<input type="radio"/> Sometimes (monthly)	Have you used a computer to communicate with foreigners?
<input type="radio"/> Rarely (semiannually)	<input type="radio"/> Yes
<input type="radio"/> I have no opinion	<input type="radio"/> No
Are you interested to interact with foreigners?	Have you used a computer to communicate with Brazillian users?
<input type="radio"/> Yes	<input type="radio"/> Yes
<input type="radio"/> No	<input type="radio"/> No

Figura 4-5 Exemplos de perguntas do questionário pré-sessão do usuário canadense

4.5.4 Questionário Pós-Sessão: Usuário Canadense

Ao final da realização da sessão de bate-papo, foi aplicado o questionário pós-sessão do usuário canadense (Apêndice D) a fim de coletar sua opinião sobre o nível de inglês do usuário brasileiro, sobre a experiência de se comunicar com alguém do Brasil, sugestões de melhorias e grau de interesse em se produzir mensagens em português. A partir das respostas deste questionário, é possível fazer uma comparação com o questionário pré-sessão do usuário brasileiro e observar o papel do C2C e de seus recursos nos resultados. A Figura 4-6 mostra algumas perguntas do questionário.

What do you think about the level of English of the Brazilian user?

- Advanced
- Intermediate
- Basic
- Other:

From 1 (min) to 10 (max), how do you evaluate the English level of the conversation?

Select an option ▼

Figura 4-6 Exemplos de perguntas do questionário pós-sessão do usuário canadense

4.6 Primeira Etapa do Estudo Piloto

Nesta primeira etapa, foram realizadas 5 (cinco) sessões de bate-papo (10 usuários no total). Ela é composta pelos procedimentos já descritos na seção 4.4, sendo que para o envio do questionário foi utilizado o Google Docs Forms¹⁸ por meio do e-mail de cada participante. O envio das questões de sugestão de conversa também foi feito por meio de e-mail, com um documento texto em anexo.

Para a gravação da tela do usuário brasileiro utilizou-se o Teamviewer¹⁹ em conjunto com o recordMyDesktop²⁰ em ambiente Linux, sendo o primeiro para

¹⁸ <http://docs.google.com/support/bin/topic.py?topic=1360904>

¹⁹ <http://www.teamviewer.com/pt/index.aspx>

²⁰ <http://recordmydesktop.sourceforge.net/about.php>

transmissão remota e o segundo para a gravação em si. Neste estudo piloto, apenas a tela do usuário brasileiro foi gravada, devido ao grau de importância da interação (já que um dos objetivos era analisar o uso do senso comum) e a disponibilidade de recursos.

O envio do questionário pós-sessão seguiu os moldes do envio do questionário pré-sessão, sendo enviado por e-mail a cada um dos participantes da sessão.

4.6.1 Respostas do Questionário Pré-Sessão: Usuário Brasileiro

A Tabela 4-1 apresenta alguns dados do perfil dos usuários brasileiros (identificados por letras maiúsculas) participantes da primeira etapa do estudo piloto. Tais dados foram retirados do cadastro dos usuários na base de dados do projeto OMCS-Br.

Tabela 4-1 Informações do cadastro dos usuários brasileiros

Usuário	Idade	Escolaridade	Cidade
A	22	Superior completo	Taubaté-SP
B	24	Superior incompleto	Santos-SP
C	22	Superior incompleto	Taubaté-SP
D	27	Superior completo	São Carlos-SP
E	24	Latu Senso	São Carlos-SP

Outros dados coletados por meio do questionário pré-sessão podem ser visualizados na Tabela 4-2.

Tabela 4-2 Perfil dos usuários brasileiros

Usuário	Profissão	Nível de Inglês	Escrita/leitura	Tempo estudando Inglês
A	Bióloga	Básico	5	3 meses
B	Estagiário	Avançado	7	5 anos
C	Engenheiro Ambiental	Intermediário	8	3 anos
D	Bacharel em Ciências	Básico	7	3 anos

	Biológicas			
E	Terapeuta Ocupacional	Básico	4	1 ano

Com relação ao uso do computador para se comunicarem com outras pessoas, todos relataram que utilizam muito frequentemente (diariamente) ferramentas como msn, chat, Skype e Livemocha. Apenas o usuário C não utilizou o computador para aprender ou estudar inglês e o usuário D nunca utilizou o computador para conversar em inglês com alguém. Alguns meios providos pelo computador para o estudo de inglês, segundo os usuários, são programas, sites de escolas de inglês, *chat*, Livemocha.

4.6.2 Respostas do Questionário Pré-Sessão: Usuário Canadense

Com relação aos usuários canadenses (identificados por números), a Tabela 4-3 apresenta alguns dados retirados do cadastro dos mesmos no C2C.

Tabela 4-3 Informações do cadastro dos usuários canadenses

Usuário	Idade	Escolaridade
1	26	Superior completo
2	29	Mestrado
3	27	Mestrado
4	24	Mestrado
5	29	Superior completo

Outros dados coletados por meio do questionário pré-sessão podem ser visualizados na Tabela 4-4.

Tabela 4-4 Perfil dos usuários canadenses

Usuário	Profissão	Usa do computador para comunicar com estrangeiros	Interesse em se comunicar com estrangeiros
1	Estudante de Mestrado	Não	Sim
2	Doutorando	Sim	Não

3	Doutorando	Não	Sim
4	Estudante de Mestrado	Sim	Sim
5	Estudante de Mestrado	Sim	Sim

Todos os usuários canadenses desta etapa relataram que fazem uso muito frequentemente (diariamente) do computador para se comunicar com outra pessoa. Entretanto, nenhum deles utilizou o computador para se comunicar com brasileiros, sendo essa sua primeira experiência.

4.6.3 Interação dos Usuários com C2C com Apoio da Tradução Cultural

Após a aplicação dos questionários, foram realizadas as sessões de bate-papo, as quais tiveram duração variando de 31 a 45 minutos. Alguns dados extraídos das gravações em vídeo da interação dos usuários brasileiros A-E são apresentados na Tabela 4-5.

Tabela 4-5 Informações da sessão de bate-papo

Usuário	Tempo da Sessão (min.)	Total de Mensagens Enviadas	Média (msg/min)	Uso do Tradutor Automático	Uso do Tradutor Cultural para Pesquisas Diretas
A	31	48	1,55	12	1
B	45	96	2,13	92	0
C	39	66	1,69	11	2
D	45	53	1,18	27	2
E	36	24	0,67	30	1

Observou-se que apenas o usuário E teve necessidade de olhar as questões de sugestão de conversa durante a sessão de bate-papo. Este usuário também foi o que produziu um menor número de mensagens por minuto. Uma justificativa para isso é esta sessão ser a primeira vez que o usuário em questão se comunica com um falante da língua inglesa, conforme relatado no questionário pré-sessão.

Os dados preliminares apontam para uma relação entre nível de inglês e quantidade média de mensagens por minuto. Nota-se que os usuários B e C, que possuem nível de inglês avançado e intermediário, respectivamente, tiveram maior média de mensagens enviadas por minuto do que os outros usuários de nível básico.

Nota-se que todos os usuários utilizaram o tradutor automático, com destaque para o usuário B que utilizou o tradutor para um alto número em comparação às mensagens enviadas. Os usuários de nível básico também utilizaram bastante o tradutor automático.

Com relação ao uso do tradutor cultural, verifica-se que o número de uso direto, isto é, busca direta no campo deste recurso foi pequeno. Durante a criação das mensagens em português, o tradutor cultural ofereceu sugestões que, no entanto, não puderam ser contabilizadas apenas na observação das gravações. No caso do usuário A, o mesmo pesquisou sobre o termo em inglês “*belated*”, ao receber uma mensagem contendo essa palavra. Os usuários C e E também utilizaram o tradutor cultural para pesquisar termos em inglês, sendo que o primeiro pesquisou pelas palavras “*swedish*” e “*Sweden*” e o segundo, pela expressão “*pop down*”. O usuário D realizou a pesquisa por palavras em português, como “congresso” e “*bioinformática*”.

Percebe-se que um problema recorrente da base de conhecimento do projeto OMCS-Br é com relação ao tamanho da base. Os principais temas que surgiram nas conversas estão relacionados à profissão, idade, região de moradia, horário de cada país, clima, comida, filme, preferências pessoais, namoro e estudo. Acredita-se que se a base do projeto OMCS-Br tivesse uma cobertura maior para esses temas, o tradutor cultural poderia trazer mais sugestões.

4.6.4 Respostas do Questionário Pós-Sessão: Usuário Brasileiro

Após cada sessão de bate-papo, foi enviado o questionário pós-sessão. A Tabela 4-6 apresenta a opinião dos usuários brasileiros com relação ao C2C e seus recursos.

Tabela 4-6 Respostas do questionário pós-sessão dos brasileiros

Usuário	Interação	Desempenho	Tradutor	Tradutor
---------	-----------	------------	----------	----------

	com C2C		Automático	Cultural
A	Difícil	Rápido	Útil	Inútil
B	Muito fácil	Rápido	Muito útil	Não opinou
C	Fácil	Indiferente	Útil	Útil
D	Muito fácil	Rápido	Muito útil	Indiferente
E	Fácil	Lento	Indiferente	Indiferente

Ao serem perguntados sobre sugestões de melhorias para o *chat*, houve duas frentes: uma no que diz respeito à ferramenta e outra no quesito de funcionalidade. Para melhorar a ferramenta, algumas sugestões foram introduzir opção de câmera, mecanismos para identificação de que o outro participante está digitando uma mensagem e adaptação para o Internet Explorer. Outra frente de sugestões se refere ao recurso de tradução: a possibilidade de tradução para outras línguas, tradução do inglês para português, melhorias no tradutor automático e aumento do banco de dados do mesmo. Em relação ao quesito tradutor automático, é interessante notar que, mesmo a maioria dos usuários brasileiros tendo avaliado o tradutor automático positivamente, eles sugerem melhorias para o mesmo.

Sobre o tradutor cultural, as respostas conferem com a análise feita na seção 4.6.3. O usuário A disse que o tradutor cultural “não ajudou em muita coisa e não contém algumas palavras”. O usuário D e E opinaram que a busca realizada não mostrou resultados. Essas opiniões indicam que o problema do tamanho da base de dados do projeto OMCS-Br afetou o uso do tradutor cultural. Mesmo assim, os usuários C e D apontaram que o recurso é uma ajuda importante. O usuário B relatou que não utilizou o tradutor cultural, pois estava entretido na conversa.

Na questão da experiência de se comunicar com um usuário canadense a partir do C2C, os usuários A e E acharam um pouco difícil por causa da falta de conhecimento da língua inglesa. O usuário B relatou que a experiência foi muito boa, pois ele pode praticar inglês e tirar algumas dúvidas. O usuário C achou tranquila a comunicação e teve algumas dúvidas nas quais precisou utilizar recursos externos já que os do C2C não ajudaram. O usuário D respondeu que a experiência foi interessante, a conversa foi amigável e que foi possível compreender as mensagens e escrevê-las com ou sem auxílio do tradutor. Ele relatou também que fazia uso de *chats* na internet e achou a proposta interessante.

4.6.5 Respostas do Questionário Pós-Sessão: Usuário Canadense

Algumas respostas do questionário pós-sessão dos usuários canadenses são apresentadas na Tabela 4-7.

Tabela 4-7 Respostas do questionário pós-sessão dos canadenses

Usuário	Interação com C2C	Desempenho	Nível de inglês do parceiro	Escrita/leitura
1	Fácil	Rápido	Básico	6
2	Muito fácil	Lento	Intermediário	5
3	Muito fácil	Muito rápido	Intermediário	7
4	Muito fácil	Muito rápido	Intermediário	7
5	Fácil	Muito rápido	Intermediário	5

Ao serem perguntadas sobre as sugestões de melhorias para o C2C, as respostas foram similares aos usuários brasileiros: adicionar vídeo e mecanismos para identificação de que o outro participante está escrevendo. Percebe-se que esta última sugestão é ainda mais importante para este tipo de usuário, pois foi relatado que em vários momentos eles não sabiam se os brasileiros estavam escrevendo ou não tinham assunto para falar. Essa constatação se deve, em grande parte, ao fato de que o usuário brasileiro leva mais tempo para produzir uma mensagem já que dispõe de recursos (tradutores cultural e automático) que demandam seu tempo e atenção.

No quesito de utilidade da adição de recursos para a produção de mensagens em português, os usuários, com exceção do usuário B, relataram que seria útil ou muito útil o *chat* ter esses recursos. O usuário B relatou que tais recursos seriam inúteis, pois ele não tem interesse no aprendizado de outra língua.

Com relação à experiência de se comunicar com um brasileiro, os usuários A, C e D disseram ter se entretido bastante. O usuário B achou o usuário brasileiro amigável, mas que a conversa foi um pouco artificial. O usuário E comentou que o parceiro estava relutante em se comunicar.

Com as respostas do questionário, percebe-se que ocorreram diferenças com relação à percepção do nível de inglês por parte de brasileiros e canadenses. A

Tabela 4-8 resume as respostas das questões relacionadas a isso. Percebe-se que usuários brasileiros que se consideravam de nível básico tiveram um conceito maior (tanto no nível ou somente na nota de escrita e leitura) por parte dos seus parceiros canadenses.

Tabela 4-8 Comparação do nível de inglês dos brasileiros

Usuário	Nível de Inglês (relatado pelo usuário brasileiro)	Nível de Inglês (relatado pelo usuário canadense)	Nota de Escrita/Leitura (relatada pelo usuário brasileiro)	Nota de Escrita/Leitura (relatada pelo usuário canadense)
A	Básico	Básico	5	6
B	Avançado	Intermediário	7	5
C	Intermediário	Intermediário	8	7
D	Básico	Intermediário	7	7
E	Básico	Intermediário	4	5

Esses resultados preliminares mostram indícios de que a primeira e a terceira hipótese do estudo piloto se confirmam, já que a comunicação entre brasileiros e canadenses por meio do C2C se mostrou satisfatória para ambos os lados. Para verificar a segunda hipótese, relacionada com o uso do senso comum por meio do tradutor cultural, foi necessário realizar uma segunda etapa do estudo piloto. Nesta etapa, o tradutor cultural foi omitido e com a comparação dos resultados de ambas as etapas pode-se observar qual o impacto desta modificação.

4.7 Segunda Etapa do Estudo Piloto

Nesta segunda etapa, foram realizadas 3 (três) sessões de bate-papo (6 usuários no total). Os mesmos procedimentos da primeira etapa se aplicam aqui. Houve apenas uma alteração quanto à tecnologia utilizada para gravação da tela do

usuário brasileiro. Foi utilizado o programa Teamviewer em conjunto com o Camtasia Studio²¹ em ambiente Windows.

4.7.1 Respostas do Questionário Pré-Sessão: Usuário Brasileiro

Os dados relacionados ao perfil dos usuários brasileiros desta etapa são apresentados na Tabela 4-9. Estes dados foram coletados por meio do cadastro dos usuários na base de dados do projeto OMCS-Br.

Tabela 4-9 Informações do cadastro dos usuários brasileiros (segunda etapa)

Usuário	Idade	Escolaridade	Cidade
F	26	Superior completo	São Carlos-SP
G	23	Superior completo	São Carlos-SP
H	23	Mestrado	São Carlos-SP

Por meio do questionário pré-sessão foram coletados alguns dados apresentados na Tabela 4-10.

Tabela 4-10 Perfil dos usuários brasileiros (segunda etapa)

Usuário	Profissão	Nível de Inglês	Escrita/leitura	Tempo estudando Inglês
F	Estudante	Intermediário	5	2 anos
G	Engenheira da Computação	Básico	4	6 meses
H	Estudante	Intermediário	5	3 anos

Todos os usuários desta etapa relataram que utilizam o computador muito frequentemente (diariamente) para se comunicar com outras pessoas e têm interesse em utilizar ferramentas que ajudem na comunicação em língua estrangeira. Apenas o usuário F nunca utilizou o computador para conversar em inglês.

Entre algumas ferramentas utilizadas para aprender inglês estão LiveMocha e Meebo. As ferramentas citadas que foram utilizadas para se comunicar com pessoas estrangeiras são o Gtalk e Msn.

²¹ <http://www.techsmith.com/camtasia/>

4.7.2 Respostas do Questionário Pré-Sessão: Usuário Canadense

O perfil dos usuários canadenses desta segunda etapa é apresentado na Tabela 4-11.

Tabela 4-11 Perfil dos usuários canadenses (segunda etapa)

Usuário	Idade	Profissão	Usa do computador para comunicar com estrangeiros
6	23	Estudante de Mestrado	Sim
7	26	Estudante de Mestrado	Não
8	29	Estudante de Mestrado	Sim

Todos os usuários desta etapa também relataram utilizar muito frequentemente o computador para se comunicar com outras pessoas. Apenas os usuários 6 e 8 já tinham utilizado o computador para conversar com estrangeiros por meio de ferramentas como Messenger e Skype. Todos tinham interesse em se comunicar com estrangeiros, entretanto nenhum tinha se comunicado com brasileiros.

4.7.3 Utilizando o C2C sem Apoio de Tradução Cultural

Após a aplicação dos questionários, foram realizadas as sessões de bate-papo que tiveram duração de 33 a 40 minutos. Alguns dados extraídos das interações gravadas são apresentados na Tabela 4-12.

Tabela 4-12 Informações das sessões de bate-papo (segunda etapa)

Usuário	Tempo da Sessão (min.)	Total de Mensagens Enviadas	Média (msg/min)	Uso do Tradutor Automático
F	33	45	1,36	24
G	37	24	0,65	25

H	40	35	0,87	15
---	----	----	------	----

Observou-se que os usuários G e H necessitaram recorrer ao documento de sugestões de conversa durante a realização da sessão de bate-papo. Pode-se perceber que os usuários que recorreram a tal documento têm uma produção de mensagens pequena em comparação aos demais. Um pequeno problema ocorreu na sessão do usuário H: o participante canadense acidentalmente fechou a janela do C2C e, ao reabrir essa janela, a interface do usuário brasileiro foi reiniciada.

Conforme também verificado na primeira etapa, observou-se que usuários com nível mais básico de inglês utilizam o tradutor automático mais vezes. Durante a sessão do usuário F, houve o uso de um tradutor automático externo (que não foi possível identificar pelas gravações) para traduções do inglês para português, ou seja, para as mensagens vindas do usuário canadense. Alguns temas recorrentes nas conversas desta etapa estão relacionados à profissão, região de moradia, clima, conhecimento sobre o Brasil e preferências pessoais.

4.7.4 Respostas do Questionário Pós-Sessão: Usuário Brasileiro

Seguindo os procedimentos do estudo piloto, após a sessão de bate-papo, foi aplicado o questionário pós-sessão. As respostas deste questionário são apresentadas na Tabela 4-13.

Tabela 4-13 Respostas do questionário pós-sessão dos brasileiros

Usuário	Interação com C2C	Desempenho	Tradutor Automático
F	Muito fácil	Muito rápido	Útil
G	Muito fácil	Rápido	Muito útil
H	Muito fácil	Rápido	Muito útil

Foi perguntado aos usuários sobre a possibilidade do C2C ter mecanismos para tradução de expressões culturais (gírias, coloquialismos etc.). O usuário F diz que seria útil, pois facilitaria a interação com a pessoa de outro país. O usuário G respondeu que seria muito útil para tornar a conversa mais informal e próxima da realidade. O usuário H relatou que seria muito útil, pois a tradução “ao pé da letra” não iria condizer nesse caso.

Segundo os usuários F e H, uma sugestão para o C2C seria a adição de um tradutor automático do inglês para o português. O usuário G gostaria de sugestões de expressões, de modo a evitar confusões por parte do outro participante.

Pelas respostas dos usuários, percebe-se que o conceito do tradutor cultural promete ser um mecanismo interessante para auxílio na criação de mensagens. Apesar de ter qualificações positivas, apenas o tradutor automático da maneira que está atualmente parece não suprir todas as necessidades de seus utilizadores.

4.7.5 Respostas do Questionário Pós-Sessão: Usuário Canadense

Algumas respostas dos usuários canadenses são apresentadas na Tabela 4-14.

Tabela 4-14 Respostas do questionário pós-sessão canadense (segunda etapa)

Usuário	Interação com C2C	Desempenho	Nível de inglês do parceiro	Escrita/leitura
6	Muito fácil	Rápido	Intermediário	7
7	Fácil	Rápido	Básico	5
8	Fácil	Indiferente	Avançado	7

Ao serem perguntadas sobre sugestões de melhorias para o C2C, o usuário 6 cogitou a possibilidade de adição de uma figura ou foto do participante a fim de aumentar a conexão entre as pessoas. O usuário 8 sugeriu a adição de uma caixa de confirmação ao se fechar a janela do *chat*, visto que ele acidentalmente fechou a janela em sua sessão de bate-papo.

Com relação à experiência de comunicação com um brasileiro, os usuários canadenses acharam ela positiva. O usuário 6 relatou também que por causa do limite de tempo e a natureza dos chat fazem com que seja difícil e artificial conhecer a outra pessoa. Apesar disso, ele acrescentou que essa é uma boa maneira de conhecer pessoas de outras culturas. O usuário 7 respondeu que o conteúdo das mensagens foram simples, mas relativamente fácil de entender. O usuário 8 disse que gostou da conversa e pareceu contente, pois conseguiu uma “receita de polenta” (comida típica brasileira).

Comparando os resultados obtidos no questionário pré-sessão dos brasileiros com as respostas do questionário pós-sessão dos canadenses, tem-se a Tabela 4-15. É possível perceber que os usuários brasileiros tiveram um conceito maior no nível de inglês ou na avaliação de escrita e leitura por parte de seus parceiros canadenses. Na próxima seção, é feita uma análise geral e comparativa das duas etapas do estudo piloto.

Tabela 4-15 Comparação do nível de inglês dos brasileiros (segunda etapa)

Usuário	Nível de Inglês (relatado pelo usuário brasileiro)	Nível de Inglês (relatado pelo usuário canadense)	Nota de Escrita/Leitura (relatada pelo usuário brasileiro)	Nota de Escrita/Leitura (relatada pelo usuário canadense)
F	Intermediário	Intermediário	5	7
G	Básico	Básico	4	5
H	Intermediário	Avançado	5	7

4.8 Análise Geral dos Resultados

Comparando-se a primeira e a segunda etapa, não há a possibilidade de se afirmar o quanto o tradutor cultural influenciou no resultado final, pois existem diversos fatores e variáveis que podem influenciar nos resultados, tais como: a timidez, interesses em comum, o quão comunicativo é cada usuário. Apesar de cada etapa possuir número de sessões diferentes, foram observados alguns indícios de que o tradutor cultural e, portanto, o senso comum impacta indiretamente na conversa entre pessoas de diferentes culturas.

Escolhendo-se as médias de envio de mensagens por minuto da primeira etapa e as médias da segunda etapa, é possível montar o gráfico apresentado na Figura 4-7.. De acordo com o gráfico, os usuários brasileiros da primeira etapa tiveram uma produção maior de mensagens enviadas do que os usuários da segunda etapa. Este resultado pode ser um indício de que as sugestões de senso comum, juntamente com suas traduções automáticas, facilitou a escrita de

mensagens, já que pode ter provido uma maior confiabilidade na tradução automática. Como exemplo, o usuário D mandou traduzir a palavra “chato” e tanto o tradutor automático quanto o tradutor cultural sugeriram a palavra “*boring*”. Além disso, com maior produção de mensagens um maior engajamento desses usuários, conforme já relatado em Silva (2009).

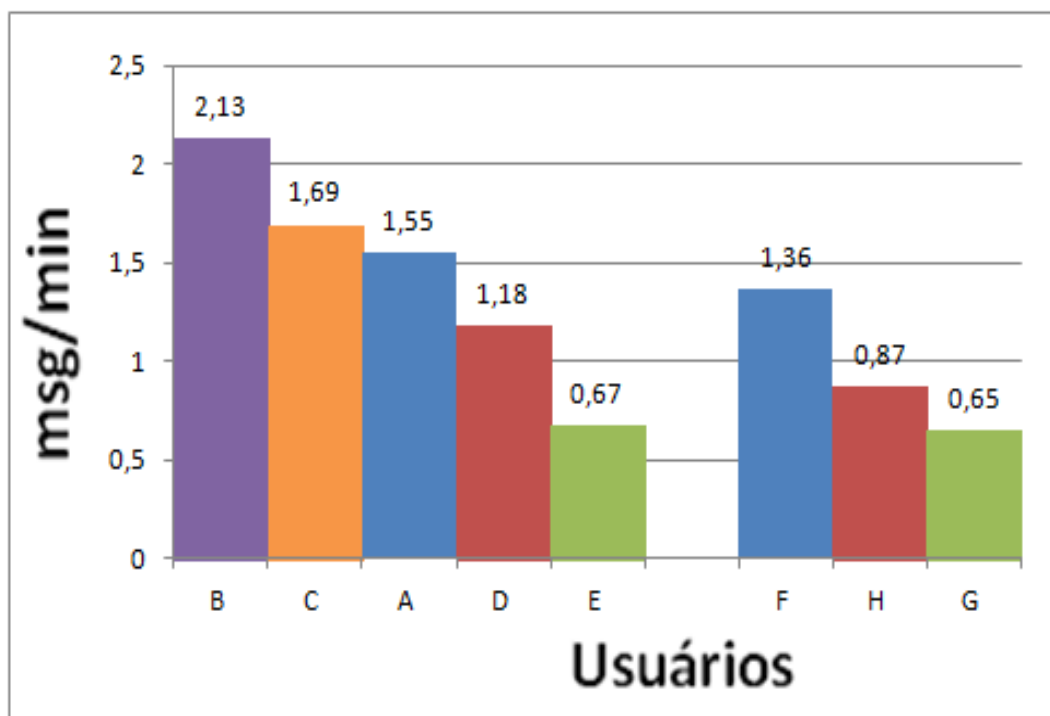


Figura 4-7 Médias de mensagens enviadas por minuto

4.9 Considerações Finais

Este capítulo teve como objetivo apresentar a forma com que o estudo piloto para a avaliação do C2C foi conduzido, assim como, descrever os resultados obtidos por meio das duas etapas definidas. O próximo capítulo traz uma síntese dos principais resultados obtidos por meio do estudo piloto e trabalhos futuros.

Capítulo 5

CONCLUSÕES E TRABALHOS FUTUROS

5.1 Síntese dos Principais Resultados

Este trabalho apresentou o Culture-to-Chat, também conhecido como C2C, um *chat* bilíngue com apoio de mecanismos de tradução automática e senso comum. Esta ferramenta tem como objetivo auxiliar os usuários na criação de mensagens em língua estrangeira de modo facilitar a comunicação entre pessoas de diferentes culturas. Até o presente momento, o C2C trabalha com os idiomas português (Brasil) e inglês.

Com os resultados obtidos por meio do estudo piloto foi possível perceber que o C2C é um meio viável para este tipo de comunicação. Uma vez que, em geral, os resultados se mostraram positivos, tanto no quesito de ferramenta como canal de comunicação quanto funcional para auxiliar o usuário na escrita de mensagens. Foi observado, por meio das análises dos artefatos do estudo piloto e da opinião dos usuários, que o C2C promove a troca de experiências referentes à cultura e idioma dos participantes.

O estudo piloto mostrou indícios de que o senso comum tem influência na comunicação de pessoas de diferentes culturas. Conforme discutido na seção 4.8, mesmo não havendo uso direto intenso do tradutor cultural, recurso do C2C que faz uso do senso comum, as sugestões apresentadas por este recurso podem ter influenciado na motivação e engajamento dos usuários na conversa.

Notou-se que a opção de pesquisa direta no tradutor cultural é requerida pelos usuários, mas que no presente momento deste trabalho a quantidade de fatos contidos na base de conhecimento do projeto OMCS-Br não é suficiente para atender as necessidades dos usuários. Este trabalho contribuiu com este mesmo projeto com a criação de *templates* para coleta de fatos relacionados a gírias, como também, estimulando os usuários da web na inserção de novos fatos na base por meio de desafios.

Outra grande contribuição para o projeto OMCS-Br é o mecanismo de retroalimentação do C2C, conforme descrito na seção 3.5. Por meio deste mecanismo, o C2C pode extrair da interação do usuário relações de Minsky e inseri-las na base de conhecimento do projeto. Esta inserção funciona como um aumento de frequência (variável f) das relações já existentes, impactando diretamente na lista de sugestões do tradutor cultural. Pode-se dizer que o C2C é um sistema que se modifica à medida que é utilizado.

Com a implementação deste trabalho, há diversos benefícios gerados para outros projetos do LIA. Um deles é a exploração do código da ConceptNet. Por meio deste trabalho, algumas funções utilizadas pelo mecanismo de processamento de língua natural do módulo gerador da ConceptNet, como o normalizador, foram disponibilizadas na API utilizada pelas aplicações para acesso a essa rede semântica. Com isso, a busca por um termo pode ser expandida, normalizando-se este antes de submetê-lo à pesquisa.

Outro benefício é a exploração da linguagem Ruby On Rails. Já visada pelo LIA, este trabalho foi o pioneiro a utilizar tal tecnologia neste laboratório e se mostrou compatível com outros mecanismos utilizados pelo projeto OMCS-Br. O protótipo web do C2C pode servir de apoio para outros trabalhos que desejam utilizar tal tecnologia.

5.2 Trabalhos Futuros

Apesar dos resultados promissores do estudo piloto, observou-se que algumas melhorias podem ser feitas para facilitar o uso do C2C. Algumas delas são descritas a seguir.

Um trabalho futuro é a investigação e adequação do C2C para utilizar outros tradutores automáticos além do Google Translate.

Considerando os dados obtidos no estudo piloto, verifica-se a necessidade da realização de um estudo mais geral e ampliado como, por exemplo, um estudo de caso.

Outra frente de trabalho é a inserção de um filtro na ConceptNet, sendo que, uma vez que a base possuísse um tamanho adequado de sugestões culturais para os usuários, haveria uma seleção das sugestões de senso comum baseadas no perfil do usuários considerando a faixa etária, gênero, escolaridade, região em que vive.

Durante as sessões de bate-papo, os usuários necessitaram de um recurso de tradução do inglês para o português. Tal recurso pode ser adicionado como uma funcionalidade extra do tradutor cultural: na procura por termos em inglês, tal termo é traduzido e sua tradução é submetida à ConceptNet para recuperação de sinônimos. Assim, o usuário teria disponível uma lista de traduções sobre determinada palavra em língua estrangeira. Este novo recurso seria outra forma de se utilizar o senso comum em traduções culturais.

A investigação de outras relações de Minsky na composição da lista de expressões do tradutor cultural é uma frente de trabalho futuro. Por exemplo, as relações “*motivationOf*”, “*desireOf*”, “*effectOf*” trabalham com os sentimentos das pessoas perante uma determinada situação. Com isso, é possível definir quais grupos de palavras apresentam um sentido mais positivo ou negativo, trabalhando então tanto com o nível semântico quanto pragmático (ALMEIDA e BARANAUSKAS, 2008). Essa nova funcionalidade pode auxiliar na escolha dos termos utilizados pelo usuário remetente.

O modelo de fluxo da informação pode ser adaptado para outras línguas. Assim, com pequenos ajustes no C2C, o mesmo passaria a suportar comunicação entre falantes do português e de outras línguas além do inglês.

Por meio dos questionários, distribuídos aos participantes do estudo piloto, foi observado que algumas melhorias para o *chat* seriam: adição de um mecanismo para verificação se o outro usuário está digitando uma mensagem, de foto (*avatar*) do participante e recursos audiovisuais e também caixa de confirmação ao se fechar a janela de conversa.

Durante a realização deste trabalho foram publicados os seguintes artigos: (SUGIYAMA, ANACLETO e CASELI 2011), (SUGIYAMA ET AL, 2010), (SUGIYAMA, ANACLETO e CASELI, 2010), (CASELI, SUGIYAMA e ANACLETO, 2010) e (FREITAS et al, 2010).

REFERÊNCIAS

AIKEN, M.; PARK, M. Enhancing bilingual electronic group meeting comprehension with round-trip translations. *Int. J. Inf. Syst. Chang. Manage.*, Inderscience Publishers, Inderscience Publishers, Geneva, SWITZERLAND, v. 4, p. 103-116, April 2009. ISSN 1479-3121. Disponível em: <<http://dx.doi.org/10.1504/IJISCM.2009.032750>>.

ALMEIDA, L. D. A.; BARANAUSKAS, M. C. C. Um prospecto de sistemas colaborativos: modelos e frameworks. In: *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*. Porto Alegre, Brazil, Brazil: Sociedade Brasileira de Computação, 2008. (IHC '08), p. 204-213. ISBN 978-85-7669-203-4. Disponível em: <<http://dl.acm.org/citation.cfm?id=1497470.1497493>>.

ANACLETO, J. et al. Can common sense uncover cultural differences in computer applications. In: *In Proc. IFIP WCC2006*, Springer-Verlag. [S.l.: s.n.], 2006. p. 1-10.

ANACLETO, J. C.; FELS, S.; VILLENA, J. M. R. Design of a web-based therapist tool to promote emotional closeness. In: *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2010. (CHI EA '10), p. 3565-3570. ISBN 978-1-60558-930-5. Disponível em: <<http://doi.acm.org/10.1145/1753846.1754019>>.

BROWN, P. F. et al. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, v. 19, p. 263-311, 1993.

BUBAS, G. Computer mediated communication theories and phenomena: Factors that influence collaboration over the internet. In: *Third CARNET Users Conference*. [S.l.: s.n.], 2001. p. 1-15.

CASELI, H. de M.; NUNES, I. A. Statistical machine translation: Little changes big impacts. *Information and Human Language Technology, Brazilian Symposium in, IEEE Computer Society, Los Alamitos, CA, USA*, v. 0, p. 63-71, 2009.

CASELI, H. de M.; SUGIYAMA, B. A.; ANACLETO, J. C. Using common sense to generate culturally contextualized machine translation. In: *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. (YIWICALA '10), p. 2431. Disponível em: <<http://dl.acm.org/citation.cfm?id=1868701.1868705>>.

CASTELLS, M. *The Network Society: A Cross-Cultural Perspective*. Edward Elgar Pub, 2005. Paperback. ISBN 1845424352. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1845424352>>.

CHEN, C. L.; RAMAN, T. V. Axsjax: a talking translation bot using google im: bringing web-2.0 applications to life. In: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A). New York, NY, USA: ACM, 2008. (W4A '08), p. 54-56. ISBN 978-1-60558-153-8. Disponível em: <<http://doi.acm.org/10.1145/1368044.1368056>>.

COOPER, A.; REIMANN, R.; CRONIN, D. About face 3: the essentials of interaction design. Wiley Pub., 2007. ISBN 9780470084113. Disponível em: <<http://books.google.com/books?id=9F7gaZKd2rYC>>.

DECEMBER, J. What is Computer-Mediated Communication? 1996. Disponível em: <<http://www.december.com/john/study/cmc/what.html>>.

DODDINGTON, G. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In: Proceedings of the second international conference on Human Language Technology Research. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002. (HLT '02), p. 138-145. Disponível em: <<http://dl.acm.org/citation.cfm?id=1289189.1289273>>.

ENGELBERG, D.; SEFFA, A. A framework for rapid mid-fidelity prototyping of web sites. In: Proceedings of the IFIP 17th World Computer Congress - TC13 Stream on Usability: Gaining a Competitive Edge. Deventer, The Netherlands, The Netherlands: Kluwer, B.V., 2002. p. 203215. ISBN 1-4020-7187-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=646869.709393>>.

FREITAS, D.; SILVA, M. A. R.; VILLENA, J. M. R.; SUGIYAMA, B. A.; ASTOLFI, G.; ANACLETO J. C.. 2010. Expanding a knowledge model representation to build culturally meaningful analogies for web applications. In IADIS International Conference WWW/Internet. [S.l.: s.n.], 2010. (WWW/Internet 2010), p. 35-41. ISBN 978-972-8939-25-0.

GARCIA, A. C.; STANDLEE, A. I.; BECHKOFF, J.; CUI, Y.. 2009. Ethnographic approaches to the internet and computer-mediated communication. Journal of Contemporary Ethnography, 2009, vol. 38, no. 1, p. 52-84.

HILL, J. D.; FLYNN, K. M. Classroom instruction that works with English language learners. Alexandria, Va. :: Association for Supervision and Curriculum Development c2006. Disponível em: <<http://site.ebrary.com/lib/wmulib/Doc?id=10141226>>.

HIRVELA, A. Computer-mediated communication in esl teacher education. ELT Journal, v. 60, n. 3, p. 233-241, 2006. Disponível em: <<http://eltj.oxfordjournals.org/content/60/3/233.abstract>>.

HUTCHINS, J. Translation technology and the translator. English, n. May, p. 8-10, 1997.

KAI, J. Cross-cultural communication. Medicine, v. 33, n. 2, p. 31 - 34, 2005. ISSN 1357-3039. Ethics and Communication skills. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1357303906001745>>.

KIM, K. J.; BONK, C. J. Cross-cultural comparisons of online collaboration. *Journal of Computer-Mediated Communication*, Wiley Online Library, v. 8, n. 1, p. 0, 2002. Disponível em: <<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2002.tb00163.x>>.

KJELDGAARD, D.; ASKEGAARD, S. The glocalization of youth culture: The global youth segment as structures of common difference. *Journal of Consumer Research*, University of Chicago, acting through its Press, v. 33, n. 2, p. 231-247, 2006. Disponível em: <<http://www.journals.uchicago.edu/doi/abs/10.1086/506304>>.

LAM, W. S. E. Second language socialization in a bilingual chat room: Global and local considerations. *Language Learning and Technology*, v. 8, p. 44-65, 2004.

LEINER, B. M.; CERF, V. G.; CLARK, D. D.; KAHN, R. E.; KLEINROCK, L.; LYNCH, D. C.; POSTEL, J.; ROBERTS, L. G.; WOLFF, S.. 2009. A brief history of the internet. *SIGCOMM Comput. Commun. Rev.*, out. 2009, ACM, New York, NY, USA, vol. 39, no. 5, p. 22-31. ISSN 0146-4833. Disponível em: <<http://dx.doi.org/10.1145/1629607-.1629613>>

LENAT, D. B. et al. Cyc: toward programs with common sense. *Commun. ACM*, ACM, New York, NY, USA, v. 33, p. 30-49, August 1990. ISSN 0001-0782. Disponível em: <<http://doi.acm.org/10.1145/79173.79176>>.

LIPSMAN, A. comScore Releases Top Online Email and Instant Messenger Applications for Brazil in December 2009. February 2010. Disponível em: <http://www.comscore.com/Press_Events/Press_Releases/2010/2/Top_Online_Email_and_Instant_Messenger_Applications_for_Brazil>.

LIU, H.; SINGH, P. Conceptnet: A practical commonsense reasoning tool-kit. *BT Technology Journal*, Kluwer Academic Publishers, Hingham, MA, USA, v. 22, p. 211-226, October 2004. ISSN 1358-3948. Disponível em: <<http://dl.acm.org/citation.cfm?id=1031314.1031373>>.

MAHMUD, J.; MATTHEWS, T.; WHITTAKER, S.; MORAN, T.; LAU, T. 2011. Topika: integrating collaborative sharing with email. In *Proceedings of the 2011 annual conference on Human factors in computing systems*. New York, NY, USA: ACM, 2011. (CHI '11), p. 3161-3164. ISBN 978-1-4503-0228-9. Disponível em: <<http://doi.acm.org/10.1145/1978942-.1979410>>

MCQUAIL, D. *McQuail's Mass Communication Theory*. SAGE Publications, 2010. ISBN 9781849202923. Disponível em: <http://books.google.com/books?id=7_eyvmCjzkoC>.

MINSKY, M. *The Society of Mind*. Pages bent. Simon & Schuster, 1988. Paperback. ISBN 0671657135. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0671657135>>.

NIELSEN, J. Try to Be a Test User Sometime. August 2011. Disponível em: <<http://www.useit.com/alertbox/being-a-test-user.html>>.

NORMAN, D. A.; DRAPER, S. W. User Centered System Design; New Perspectives on Human-Computer Interaction. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1986. ISBN 0898597811.

PAPINENI, K. et al. Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. (ACL '02), p. 311-318. Disponível em: <<http://dx.doi.org/10.3115/1073083.1073135>>.

PARAMSKAS, D. M. The shape of computer-mediated communication. CALL: Media, Design and Applications, Cameron, K.C., p. 13-34, 1999.

PIMENTEL, C. Vc tb escreve axim? uma análise do léxico nos blogs de adolescentes. Revista Querubim, v. 01, n. 02, p. 84-94, 2006. Disponível em: <www.lologia.org.br/ixcnlf/4/11.htm>.

ROMISZOWSKI, A.; MASON, R. Computer mediated communication. D. H. Jonassen, 2004. (Handbook of research for educational communications and technology). ISBN 9780877782438. Disponível em: <http://books.google.com/books?id=S_9kbzLw1HYC>.

SILVA, M. A. R.; ANACLETO, J. C. Promoting collaboration through a culturally contextualized narrative game. In: FILIPE, J. et al. (Ed.). Enterprise Information Systems. [S.l.]: Springer Berlin Heidelberg, 2009, (Lecture Notes in Business Information Processing, v. 24). p. 870881. ISBN 978-3-642-01347-8.

SINGH, P. The public acquisition of commonsense knowledge. 2001. Disponível em: <citeseer.ist.psu.edu/singh02public.html>.

SNYDER, C. Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces (Interactive Technologies). 1. ed. Morgan Kaufmann, 2003. Paperback. ISBN 9781558608702. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1558608702>>.

SUGIYAMA, B. A.; ANACLETO, J. C.; CASELI, H. M. 2010. Um chat bilíngue para aprendizado da segunda língua considerando a cultura dos participantes. In 21 Simpósio Brasileiro de Informática na Educação. [S.l.: s.n.], 2010. (SBIE 2010), p. 1-4.

SUGIYAMA, B. A.; ANACLETO, J. C.; CASELI, H. M. 2011. Assisting users in a cross-cultural communication by providing culturally contextualized translations. In Proceedings of the 29th ACM international conference on Design of communication. ACM, 2011. (SIGDOC '11), p. 189194 . ISBN 978-1-4503-0936-3. Disponível em: <http://doi.acm.org/10.1145/2038476-2038513>

SUGIYAMA, B. A.; ANACLETO, J. C.; FELS, S.; CASELI, H. M. 2010. Using cultural knowledge to assist communication between people with different cultural background. In Proceedings of the 28th ACM International Conference on Design of

Communication. ACM, 2010. (SIGDOC '10), p. 183-190. ISBN 978-1-4503-0403-0. Disponível em: < <http://doi.acm.org/10.1145/1878450-1878481> >
<http://doi.acm.org/10.1145/1878450.1878481>

TENG, X. Negotiation of meaning in synchronous computer-mediated communication (CMC): The role of online chat in second language vocabulary development. Tese (Master's Thesis) Iowa State University, 2010.

TSUTSUMI, M. Uso de senso comum na detecção das diferenças culturais no contexto do projeto Open Mind Common Sense. 2006. 118 f. Dissertação (Mestrado em Ciência da Computação) – DC, UFSCar, São Carlos

TYLOR, E. Primitive culture: researches into the development of mythology, philosophy, religion, art, and custom. J. Murray, 1871. (Primitive Culture: Researches Into the Development of Mythology, Philosophy, Religion, Art, and Custom, v. 1). Disponível em: <<http://books.google.com/books?id=AucLAAAIAAJ>>.

WARSCHAUER, M. Computer-mediated collaborative learning: Theory and practice. The Modern Language Journal, Blackwell Publishing on behalf of the National Federation of Modern Language Teachers Associations, v. 81, n. 4, p. pp. 470-481, 1997. ISSN 00267902. Disponível em: <<http://www.jstor.org/stable/328890>>.

WHITTAKER, S. Supporting collaborative task management in e-mail. Hum.-Comput. Interact., L. Erlbaum Associates Inc., Hillsdale, NJ, USA, v. 20, p. 49-88, June 2005. ISSN 0737-0024.

YAMASHITA, N.; ISHIDA, T. Effects of machine translation on collaborative work. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. New York, NY, USA: ACM, 2006. (CSCW '06), p. 515-524. ISBN 1-59593-249-6. Disponível em: <<http://doi.acm.org/10.1145/1180875.1180955>>.

Apêndice A

QUESTIONÁRIO PRÉ-SESSÃO DO USUÁRIO BRASILEIRO

Experimento sobre o C2C

Questionário 1: Este formulário contém questões que ajudarão os pesquisadores do C2C a entender o perfil do usuário.

Qual sua profissão?

Você estuda ou já estudou Inglês?

- Sim
 Não

Se sim, por quanto tempo aproximadamente?

Como você considera o seu nível de Inglês?

- Avançado
 Intermediário
 Básico
 Other:

Com relação a escrita e leitura em inglês, que nota você daria para si?

Você usa o computador para se comunicar (msn, salas de bate-papo etc)?

- Muito frequentemente (diariamente)
 Frequentemente (semanalmente)
 Indiferente
 Poucas vezes (mensalmente)
 Raramente (semestralmente)
 Não tenho como opinar

Já utilizou o computador para aprender/estudar Inglês?

- Sim
 Não

Se sim, por quais meios (programa, site, chat)?

Já utilizou o computador para conversar em inglês com alguém?

- Sim
 Não

Se sim, por quais meios (programa, site, chat)?

Tem interesse em utilizar ferramentas que ajudem na comunicação em língua estrangeira?

- Sim
 Não

Apêndice B

QUESTIONÁRIO PÓS-SESSÃO DO USUÁRIO BRASILEIRO

Experimento sobre o C2C - parte 2

Questionário 2: Conte-nos como foi a experiência com o C2C.

Como foi a interação com o C2C

- Muito fácil
- Fácil
- Indiferente
- Díficil
- Muito difícil
- Não posso opinar

O que você achou do desempenho/velocidade do C2C?

- Muito rápido
- Rápido
- Indiferente
- Lento
- Muito lento
- Não posso opinar

O que você achou do tradutor automático no chat?

- Muito útil
- Útil
- Indiferente
- Inútil
- Muito inútil
- Não posso opinar

O que você achou das sugestões do tradutor cultural?

- Muito útil
- Útil
- Indiferente
- Inútil
- Muito inútil
- Não posso opinar

Por quê?

Descreva brevemente como foi a experiência de se comunicar com um estrangeiro pelo C2C

Qual(is) sugestão(ões) você daria para o chat?

Apêndice C

QUESTIONÁRIO PRÉ-SESSÃO DO USUÁRIO CANADENSE

C2C Experiment

Questionnaire 1: By completing this questionnaire, you are helping our researches to understand the C2C users' profile.

What is your profession?

Do you use a computer to interact with other people?

- Always (daily)
- Usually (weekely)
- Indifferent
- Sometimes (monthly)
- Rarely (semiannually)
- I have no opinion

Are you interested to interact with foreigners?

- Yes
- No

Have you used a computer to communicate with foreigners?

- Yes
- No

If you answered yes, what tool(s) did you used (software, site, chat)?

Have you used a computer to communicate with Brazillian users?

- Yes
- No

If you answered yes, describe briefly your experience (which tools were used, what language was used, how did you feel about it)

Apêndice D

QUESTIONÁRIO PÓS-SESSÃO DO USUÁRIO CANADENSE

C2C Experiment - Part 2

Questionnaire 2: Let us know how was your experience using the C2C.

How was the interaction with C2C?

- Very easy
- Easy
- Indifferent
- Hard
- Very hard
- I cannot opine

What do you think about the performance of C2C?

- Very fast
- Fast
- Indifferent
- Slow
- Very slow
- I cannot opine

What do you think about the level of English of the Brazillian user?

- Advanced
- Intermediate
- Basic
- Other:

From 1 (min) to 10 (max), how do you evaluate the English level of the conversation?

Select an option ▾

Briefly, how was your experience with the Brazillian user (what do you think about her/him, what do you thing about the content of the messages)?

What would you think if C2C had resources to help you write messages in Portuguese (for example, a machine translator)?

- Very useful
- Useful
- Indifferent
- Useless
- Very useless
- I cannot opine

What would you think if C2C helped you identify slang or colloquialism and provided suggestion for translation?

- Very useful
- Useful
- Indifferent
- Useless
- Very useless
- I cannot opine

What suggestions do you have for C2C?

Apêndice E

SUGESTÕES DE CONVERSA

C2C

Vamos conhecer seu colega? Abaixo estão algumas questões que ajudarão a iniciar uma conversa. Fique à vontade para utilizá-las ou não. Elabore seus próprios temas e perguntas

1- Apresente-se (Qual o seu nome? O que você faz? Onde você mora?)

2- O que você sabe sobre o país do seu colega?

3- Fale sobre o seu país para seu colega

4 - Fale sobre as pessoas do seu país (o que elas gostam, o que não gostam).

5- O que você gosta de fazer no seu tempo livre?

6- Quais feriados são mais populares no seu país

7- Como são as festas/celebrações em seu país?

8- Que tipos de presente você gosta de receber?

9- Se você tivesse bastante dinheiro, o que faria?

10- Se você fosse para o país de seu colega, o que gostaria de fazer lá