



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Guilherme Lucas de Souza

**As interações entre jogadores de jogos virtuais em fóruns *online*: uma
análise do léxico mesclado de inglês-português**

São José do Rio Preto
2021

Guilherme Lucas de Souza

As interações entre jogadores de jogos virtuais em fóruns *online*: uma análise do léxico mesclado de inglês-português

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Orientadora: Profa. Dra. Paula Tavares Pinto
Co-orientadora: Profa. Dra. Talita Serpa

São José do Rio Preto
2021

S729i	Souza, Guilherme Lucas de As interações entre jogadores de jogos virtuais em fóruns online : uma análise do léxico mesclado de inglês-português / Guilherme Lucas de Souza. -- São José do Rio Preto, 2020 78 f. : il., tabs., fotos Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto Orientadora: Profª Drª Paula Tavares Pinto Coorientadora: Profª Drª Talita Serpa 1. Linguística de Corpus. 2. Línguas em Contato. 3. Empréstimos. 4. Linguagem virtual. 5. Code-mixing. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Guilherme Lucas de Souza

As interações entre jogadores de jogos virtuais em fóruns *online*: uma análise do léxico mesclado de inglês-português

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Comissão Examinadora

Profa. Dra. Paula Tavares Pinto
UNESP – Câmpus de São José do Rio Preto
Orientador

Profa. Dra. Talita Serpa
UNESP – Câmpus de São José do Rio Preto
Co-orientador

Prof. Dr. Celso Fernando
UNESP – São José do Rio Preto

Prof. Dr. Eduardo Batista da Silva
UEG – Câmpus de Goiânia

São José do Rio Preto
13 de novembro de 2020

Dedico este trabalho a todos aqueles que me apoiaram durante este árduo processo, que entenderam minha ausência e, mesmo assim, me incentivaram a sempre continuar a busca por conhecimento. Dedico este trabalho à minha família, namorada, família e orientadora, que incansavelmente me apoiaram e me deram a oportunidade de continuar sonhando.

Àqueles que, apesar de todas as adversidades, sempre acreditaram que eu seria capaz de atingir meus objetivos.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Vanzorico e Fátima, por todo o apoio, amor, ajuda ao longo de minha vida, sempre zelando pelo meu crescimento pessoal e profissional, fazendo diversos sacrifícios e passando por privações para que pudessem prover, a mim e meus irmãos, educação e segurança, sendo exemplos de bom caráter e ética nos quais sempre me espelharei.

Às minhas avós, Edna e Tereza, por sempre estarem por perto, me acolhendo e me dando suporte desde o meu nascimento.

Aos meus irmãos, Robson, Ângela, Ana e Vitor, dedico este trabalho, por mais problemas que possamos ter (irmãos são assim mesmo), nunca deixarei de admirá-los e de ajudá-los, como sempre fizeram por mim.

Agradeço imensamente à minha namorada, Fernanda, por me acolher, me dar suporte e entender os momentos em que não pude estar presente, por saber me oferecer paz quando era preciso e incentivo quando eu menos acreditei ser possível.

Dedico este trabalho à minha orientadora, Profa. Dra. Paula Tavares Pinto, por todo o acolhimento, por todos os anos de extremo zelo e acompanhamento, pelas horas que passamos juntos em nossas reuniões tão importantes para o desenvolvimento deste trabalho, bem como para o meu desenvolvimento pessoal. Só pude chegar tão longe, pois você acreditou e me fez acreditar.

Agradeço à minha co-orientadora, Profa. Dra. Talita Serpa, pelas relevantes discussões, leituras e valiosas sugestões durante a segunda fase da pesquisa.

Aos meus amigos, Valdo, Reinaldo, Alan, Tiago e Henrique, que sempre estiveram por perto para me ajudar quando preciso, lembro-me das infinitas horas que passamos jogando RPG e que me inspiraram a imaginar e escrever.

Aos meus amigos de trabalho com os quais sempre pude contar, que me permitiram crescer dentro da minha área, me oferecendo a oportunidade de ir além como profissional das Letras.

Dedico a todos os professores que passaram pela minha vida: aos do Ensino Fundamental, que me ensinaram a reconhecer o mundo por meio das letras, aos do Ensino Médio, que me fizeram ir além, dando-me as bases para ser um cidadão crítico e, por fim, aos meus queridos professores e amigos que cultivei durante os intermináveis anos da graduação, que, para além dos ensinamentos acadêmicos, me ajudaram a ser, de fato, cidadão dono de meu próprio destino.

Ao Instituto de Biociências, Letras e Ciências Exatas (IBILCE) e todos seus funcionários por, direta e indiretamente, agirem em prol da pesquisa e oferecerem apoio acadêmico a tantos alunos como eu.

RESUMO

Atualmente, observamos a popularização de jogos eletrônicos, fomentando o surgimento e crescimento de comunidades virtuais que se utilizam de empréstimos e adaptações entre a linguagem do jogo e a língua materna de seus jogadores. Dessa forma, em nosso trabalho, tencionamos apresentar como o uso dos *MMORPG (Massive Multiplayer Online Role Playing Game)*, ou jogo *online* de interpretação em massa para múltiplos jogadores, possui características linguísticas que lhe são próprias no que tange a seus ambientes de comunicação bilíngue. Para exemplificar esses ambientes comunicativos, realizamos a coleta e a compilação de um *corpus* que contém interações entre jogadores de *League of Legends®* em um fórum *online* no Brasil. Nesse sentido, selecionamos postagens disponibilizadas pela produtora do jogo eletrônico no *website forum.br.leagueoflegends.com*, coletadas entre novembro de 2018 e março de 2019. Os resultados obtidos mostram que existe uma frequência de uso de certas palavras, as quais promovem alternâncias e misturas de inglês e português, conhecidas como *code-mixing*. Tratam-se de palavras de origem estrangeira que são inseridas em orações normalmente produzidas em língua materna (L1), sem que haja alteração morfológica ou gramatical, como, por exemplo, em nosso *corpus*, as palavras *jungle*, *lane* e *farm* (POPLACK, 1980, 1982; POPLACK; SANKOFF, 2004; APPEL; MUYSKEN, 2005; MUYSKEN, 2001; GARDNER-CHLOROS, 2009), bem como de empréstimos linguísticos adaptados, ou seja, palavras que sofrem interferências e alterações morfológicas e semânticas, como *wardar* e *farmar* (PFAFF, 1979; HOFFER, 2002; POPLACK, 1982). Dentro do *corpus* coletado, existe o uso recorrente do primeiro fenômeno, com 863 palavras em comparação ao segundo, com 409 palavras, demonstrando, assim, que o uso de vocábulos que são provenientes da língua inglesa e configuram *code-mixing* é maior. Outro aspecto observado como resultado da pesquisa reflete-se nos agrupamentos lexicais, ou seja, nas combinações de palavras que são recorrentes. Vemos que, entre os *code-mixing*, há uma adaptação no uso de artigos e preposições, como “a kill” e “a jungle”, enquanto, no que se refere aos empréstimos adaptados, as locuções verbais são presentes e recorrentes. Por fim, notamos que a linguagem observada nestes fóruns apresenta influência da língua portuguesa na produção de língua inglesa, de modo que as regras gramaticais do português levam o produtor a implementá-las nas palavras estrangeiras que são inseridas nos fóruns.

Palavras-chave: Língua inglesa. Linguística de *Corpus*. *Code-mixing*. Empréstimos Linguísticos.

ABSTRACT

Nowadays we see the popularization of electronic games, leading to the growth of virtual communities which use loans and adaptations between the players' mother tongue and the language of the game. Thus, our work aims to demonstrate how the usage of *MMORPG* (*Massive Multiplayer Online Role Playing Game*) has features that are unique according to the bilingual communication environment. To exemplify such environments, we collected and compiled a *corpus* that contains interactions between Brazilian *League of Legends*® players in an online forum. In this sense, we selected posts provided by the game developer on the website *forum.br.leagueoflegends.com*, collected from November 2018 to March 2019. The results show that there is a usage frequency of certain words, which promote shifts and mixtures between Portuguese and English, known as code-mixing. They mean that words from a foreign language are inserted into statements normally produced in the mother tongue (L1) with no morphological or grammatical changes, for example, the words “jungle”, “lane” and “farm” in our *corpus* (POPLACK, 1980, 1982; POPLACK; SANKOFF, 2004; APPEL; MUYSKEN, 2005; MUYSKEN, 2001; GARDNER-CHLOROS, 2009) as well as adapted linguistic borrowing, namely, words that suffer morphological and semantic interference, for instance, “wardar” and “farmar” (PFAFF, 1979; HOFFER, 2002; POPLACK, 1982). The data show a recurrent usage of the first phenomenon, with 863 words, in comparison with the second one, with 409 words, which demonstrates that the use of English words is more frequent. Another aspect observed as a result concerns the lexical clusters, in other words, word combinations that are frequent. We see that among the code-mixing words there is an adaptation in the usage of articles and prepositions, such as “a kill” and “a jungle”, while among the adapted borrowing ones, the verbal phrases are present and recurrent. Finally, we noticed that the language observed in the forums shows the Portuguese language influences the use of English, in the sense that Portuguese grammar rules lead the producer to implement them into the foreign words that are inserted in the forum.

Keywords: English. *Corpus* Linguistics. Code-mixing. Borrowing.

LISTA DE ILUSTRAÇÕES

Figura 1: Mapa do jogo <i>League of Legends</i> ®	13
Figura 2: <i>Layout</i> do fórum <i>online</i> do jogo <i>LoL</i> ®	14
Figura 3: Seções de fóruns no website de <i>LoL</i> ®	15
Figura 4: Postagem em fórum	16
Figura 5: <i>Layout</i> inicial do programa <i>AntConc</i> ®	37
Figura 6: Ferramenta <i>Wordlist</i> do programa <i>AntConc</i> ®	38
Figura 7: Ferramenta <i>Concordance</i> no programa <i>AntConc</i> ®	39
Figura 8: Ferramenta <i>Collocates</i> no programa <i>AntConc</i> ®	40
Figura 9: A ferramenta <i>Clusters/N-grams</i> no programa <i>AntConc</i> ®	41
Figura 10: <i>Wordlist</i> pelo programa <i>AntConc</i> ®	42
Figura 11: <i>Concordance</i> do programa <i>AntConc</i> ®	43
Figura 12: Lista de variações (itens lematizados)	44
Figura 13: Lista de palavras (<i>wordlist</i>) no programa <i>AntConc</i> ® com itens lematizados	44
Figura 14: Frequência de empréstimos e <i>code-mixing</i> no corpus de transcrições de jogadores de MMORPG	47
Figura 15: Ferramenta <i>Clusters/N-grams</i>	49
Figura 16: Exemplo de combinações utilizando a ferramenta <i>Concordance</i>	50
Figura 17: Artigos definidos + <i>code-mixing</i>	52
Figura 18: Artigo indefinido + <i>code-mixing</i>	52
Figura 19: Locução verbal envolvendo empréstimo adaptado com sufixação -ar	54
Figura 20: <i>CM</i> como objeto da ação	55
Figura 21: Contexto de uso	61
Figura 22: <i>Boxplot code-mixing</i>	62
Figura 23: <i>Boxplot</i> de empréstimos adaptados	64
Quadro 1: Informações sobre o corpus	31
Quadro 2: Tamanho dos <i>corpora</i> e sua classificação	32
Quadro 3: <i>Tokens</i> mais frequentes de <i>code-mixing</i>	50
Quadro 4: Frequência de uso de artigos definidos e indefinidos com <i>code-mixing</i>	53
Quadro 5: Verbos recorrentes e suas ligações com os <i>CM</i> e empréstimos	56
Quadro 6: Quantidade e tipologia dos <i>tokens</i> observadas no corpus	57

Quadro 7: O uso de <i>code-mixing</i> e empréstimos no discurso dos jogadores	58
Quadro 8: Frequência de uso do <i>code-mixing ward</i>	59
Quadro 9: Tipologia e exemplos	59
Quadro 10: <i>Code-mixing</i> extrapolados	63
Quadro 11: Empréstimos extrapolados.....	65

LISTA DE ABREVIATURAS E SIGLAS

CS	Code-switching
CM	Code-mixing
ESA	Entertainment Software Association
ESPM	Escola Superior de Propaganda e Marketing
LoL	League of Legends
MOBA	Multiplayer Online Battle Arena
MMORPG	Massive Multiplayer Online Role-Playing Games
RPG	Role-Playing Games

SUMÁRIO

1 INTRODUÇÃO	12
1.2 O jogo <i>League of Legends</i> ®	12
1.3 O fórum online de <i>League of Legends</i> ®	14
1.4 O que se encontra em termos linguísticos nas interações dos fóruns	16
1.5 Organização do trabalho	19
2 FUNDAMENTAÇÃO TEÓRICA	20
2.1 Os estudos das alternâncias e misturas de códigos linguísticos	20
2.2 O Code-switching	21
2.3 O <i>Code-Mixing</i>	22
2.3.1 Inserção	24
2.3.2 Alternância	26
2.3.3 Lexicalização Congruente	27
2.4 Empréstimos adaptados	30
2.5 Pressupostos da Linguística de <i>Corpus</i>	30
2.5.1 <i>Corpora</i> pequenos	33
3 MATERIAL E MÉTODO	34
3.1 Material para construção do <i>corpus</i>	34
3.2 Critérios para coleta de material	34
3.3 Procedimentos de coleta do material	35
3.4 O programa <i>AntConc</i> ®.	36
3.4.1 Avaliação das palavras pela ferramenta <i>Wordlist</i> (Lista de Palavras)	41
4 ANÁLISE E RESULTADO	46
4.1 A análise do <i>corpus</i>	46
4.1.1 Análise da frequência dos usos de <i>CM</i> e empréstimos	47
4.2 Frequência dos agrupamentos lexicais	48
4.2.1 Agrupamentos lexicais (<i>clusters</i>)	48
4.2.2 Artigo + <i>code-mixing</i> com função de substantivo	51
4.2.3 Verbos + <i>code-mixing</i> e empréstimos	54
4.3 Resultados e discussão	57
5 CONCLUSÃO E ENCAMINHAMENTOS	66
REFERÊNCIAS	69
APÊNDICE A - Frequência dos code-mixing observados no corpus	72
APÊNDICE B - Frequência dos empréstimos adaptados observados no corpus	77

1 INTRODUÇÃO

O mundo do entretenimento virtual, principalmente referente aos *videogames*, é, atualmente, um dos mercados mundiais mais importantes e proeminentes. O que se iniciou em meados do século XX, com plataformas de jogos eletrônicos rudimentares e pouco investimento, conta agora, no século XXI, com os ganhos que superam a receita da indústria cinematográfica norte-americana desde o ano de 2006, segundo a *Entertainment Software Association (ESA)*, atingindo, no ano de 2017, a arrecadação de 36 bilhões de dólares.

No Brasil não é diferente. Seguimos as mesmas projeções de crescimento em relação ao consumo de jogos eletrônicos. Em pesquisa realizada pelas empresas *Sioux Group*, *Blend New Research* e *ESPM*, notamos que 75,5% dos brasileiros têm acesso a jogos eletrônicos em qualquer plataforma de jogo. No entanto, existe um grande apreço desse público pelos jogos *mobile* (43,6% das respostas), ou seja, os que podem ser jogados pelos *smartphones*. A pesquisa aponta também que 35,2% do grupo de jogadores têm idade entre 25 e 34 anos (IGN, 2018). Uma dessas plataformas de jogos eletrônicos é a *Multiplayer Online Battle Arena (MOBA)*, que apresenta misturas de jogos de interpretação de personagens, mais conhecido como *Role-Playing Games (RPG)* e jogos de estratégia, pois o jogador precisa jogar em equipe e possuir estratégias para vencer.

Nesse *game*, o jogador cria um personagem de acordo com a possibilidade do mundo virtual, desenvolvendo-o dentro deste ambiente virtual. Diferentemente dos *Massive Multiplayer Online Role-Playing Games (MMORPG)*, que possuem milhões de jogadores interagindo em um mundo aberto e sem fim, o MOBA seria uma modalidade de jogo que permite partidas rápidas. Um dos exemplos mais famosos nesse tipo de plataforma é o jogo *League of Legends®*, doravante *LoL®* (RIOT, 2020), o mais popular e mundialmente conhecido deste gênero e que foi escolhido para esta pesquisa devido, justamente, à sua notoriedade e facilidade de acesso.

Esse ambiente de jogos eletrônicos *online* permite que formas de linguagem diferentes possam emergir nesse contexto comunicacional permeado por interações entre pessoas de diversos países, apresentando estruturas próprias, mesclando palavras e ocasionando fenômenos que poderiam ocorrer entre fronteiras geográficas.

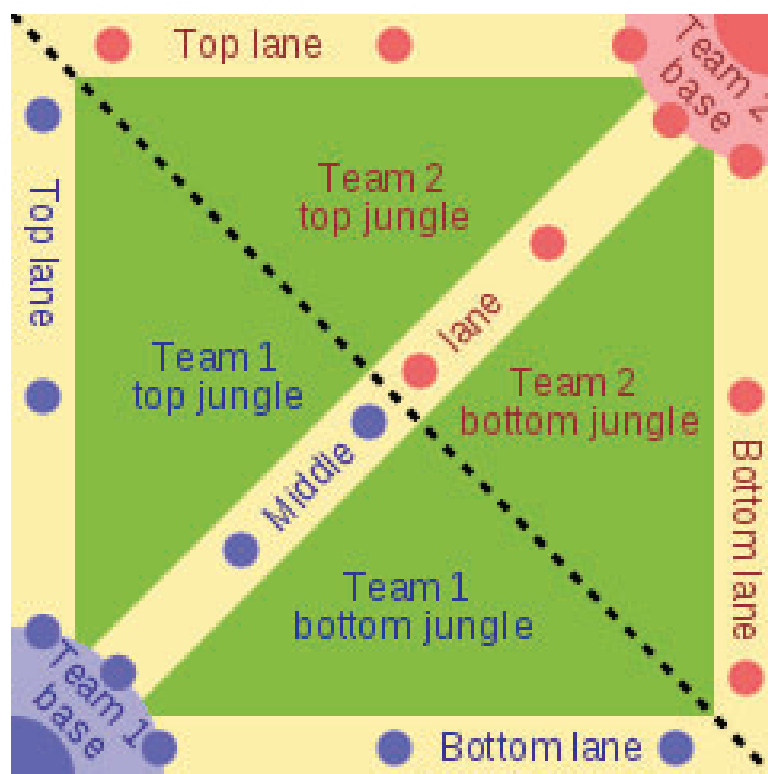
1.2 O jogo *League of Legends®*

O jogo *League of Legends* surgiu no início da década de 2010, disponibilizado pela produtora Riot Games, como um RPG de arena gratuito, conhecido como MOBA. Logo

tornou-se um dos passatempos mais populares da atualidade e, em 2012, alcançou a marca de jogo mais jogado da América do Norte e Europa, vindo a ser utilizado, até 2014, por mais de 67 milhões de pessoas (WIKIPEDIA, 2020). Atualmente, devido à sua popularidade, são realizados torneios mundiais, nos quais equipes disputam prêmios milionários.

LoL® trata-se de um jogo eletrônico que deve ser jogado por equipes. Cada equipe se organiza da melhor maneira que puder e cada jogador escolhe um de seus “campeões” ou *champions*. Uma equipe deve vencer a outra, destruindo as torres pelo caminho (*lanes*), dentro de um mapa, conhecido entre os jogadores como *jungle*, para que, assim, se chegue ao ponto de origem dos “campeões”, ou seja, dos personagens utilizados no jogo. Abaixo, podemos observar a forma como o mapa do jogo é constituído:

Figura 1: Mapa do jogo *League of Legends*®



Fonte: *League of Legends*®. (WIKIPEDIA, 2020)

Em uma partida, os jogadores se comunicam via *chat* ou microfone. Sendo assim, o uso do microfone e fones de ouvido é o mais recorrente por ser mais dinâmico, entretanto, esses aparelhos fazem a linguagem escrita ser, muitas vezes, relegada a poucas partidas ou momentos muito específicos, tornando-se um dos problemas da coleta de dados desta pesquisa, o qual será exemplificado na seção Materiais e Métodos.

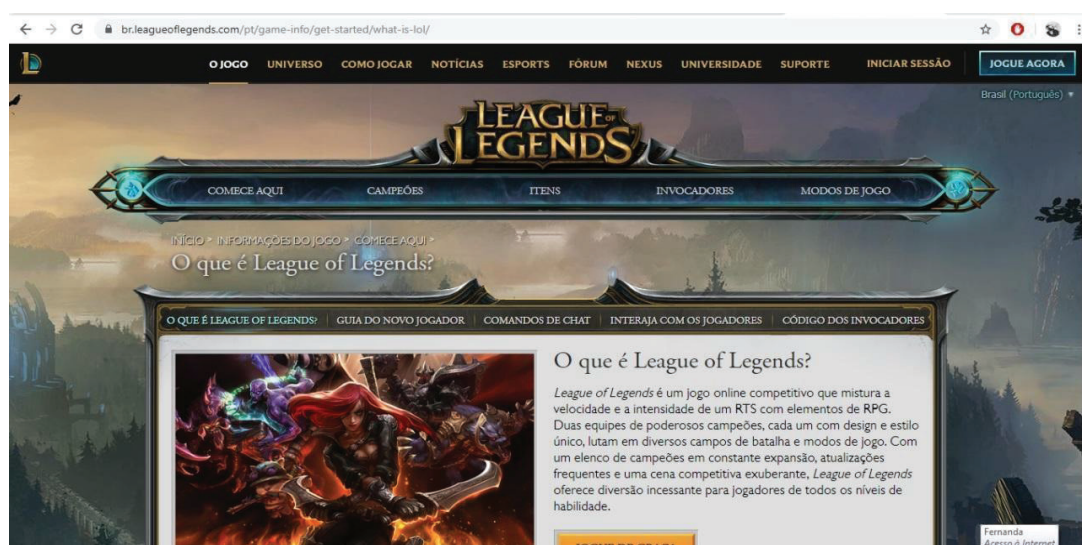
Vemos que, nessas comunicações, os jogadores utilizam diversas misturas entre as línguas portuguesa e inglesa, causando estranhamentos às pessoas que não estão inseridas nesse contexto de uso. Assim, surgiu o interesse que motivou esta pesquisa, ou seja, o de compilar esse vocabulário misturado entre a língua inglesa e portuguesa e procurar categorizá-lo, de modo a compreender seus usos em contexto.

Para termos acesso aos jogadores e à linguagem empregada, nos atemos a pesquisar e coletar informações contidas em um fórum *online*, no qual os jogadores brasileiros fazem postagens diariamente. Este fórum é gratuito e disponibilizado pela produtora do jogo¹, sendo utilizado apenas para interação entre os jogadores e a empresa, pois é nele que a produtora faz avisos, apresenta novidades, etc.

1.3 O fórum online de League of Legends®

O fórum de *LoL*® encontra-se disponível *online*, com sua versão voltada apenas aos jogadores brasileiros, e possui *layout* e textos oferecidos pela produtora em língua portuguesa. O fórum *online* pode ser acessado pelo link <https://forum.br.leagueoflegends.com/> e conta com atualizações diárias oferecidas pela produtora do jogo. Abaixo, podemos ver a figura 2, apresentando a página inicial do fórum:

Figura 2: *Layout* do fórum online do jogo *LoL*®



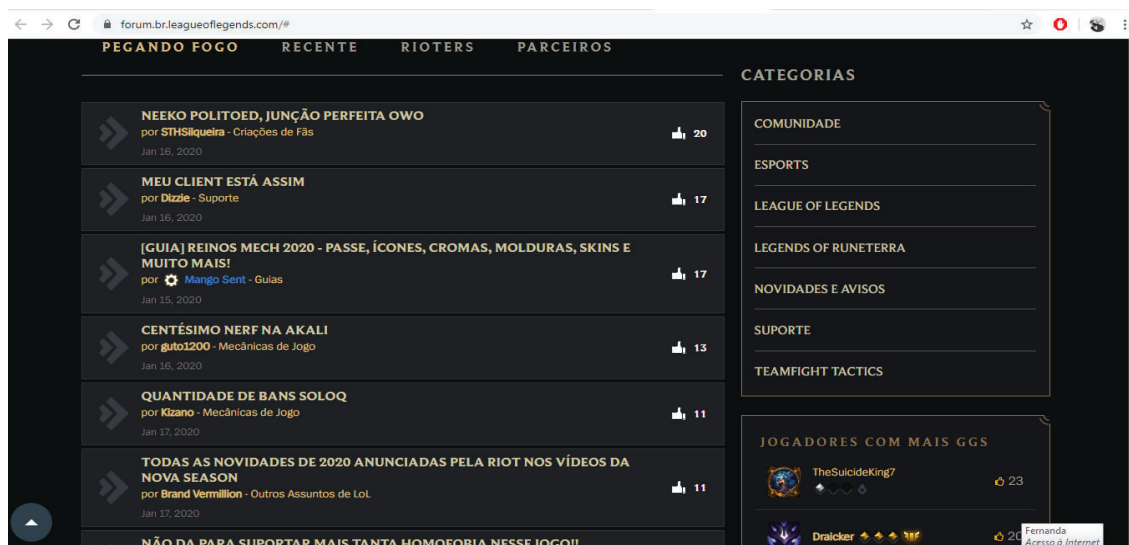
Fonte: Imagem retirada do website de *League of Legends*® (RIOT, 2019)

¹ Abordaremos sobre o fórum do jogo *League of Legends*® no próximo tópico, apresentando as características e informações encontradas no *website*.

Por meio do *website*, a empresa posta novidades, atualizações, problemas que estão ocorrendo, notícias etc. Nesse mesmo *website*, podemos fazer o *download* do jogo e saber sobre sua mecânica, modos de jogo e configurações; nele, também, podemos ter a interação de jogadores novos e veteranos, trocando informações por postagens de forma escrita no fórum. Os fóruns são divididos em:

- 1) **Comunidade**, no qual os jogadores compartilham suas informações;
 - 2) **E-sports**, no qual são divulgados resultados das ligas de e-sport do *LoL*®, por exemplo, sobre os campeonatos mundiais;
 - 3) **League of Legends**, no qual constam informações sobre o jogo, como atualizações e problemas técnicos, por exemplo;
 - 4) **Legends of Runterra**, versão *mobile* de *LoL*®, para ser jogada no celular, no qual os jogadores têm acesso às informações sobre o jogo;
 - 5) **Novidades e avisos**, no qual as postagens sobre atualizações, novidades e avisos sobre problemas são direcionadas para os jogadores;
 - 6) **Suporte**, no qual os desenvolvedores, suporte técnico e jogadores postam sobre problemas no jogo e como solucioná-los, e
 - 7) **Teamfight Tactics**, outro jogo de versão *mobile*, que, diferentemente dos outros dois, é jogado em turnos. Portanto, possui uma mecânica de jogo diferente da do *LoL*®.
- Abaixo, segue a figura 3, que apresenta a disposição dos temas dos fóruns no *website*:

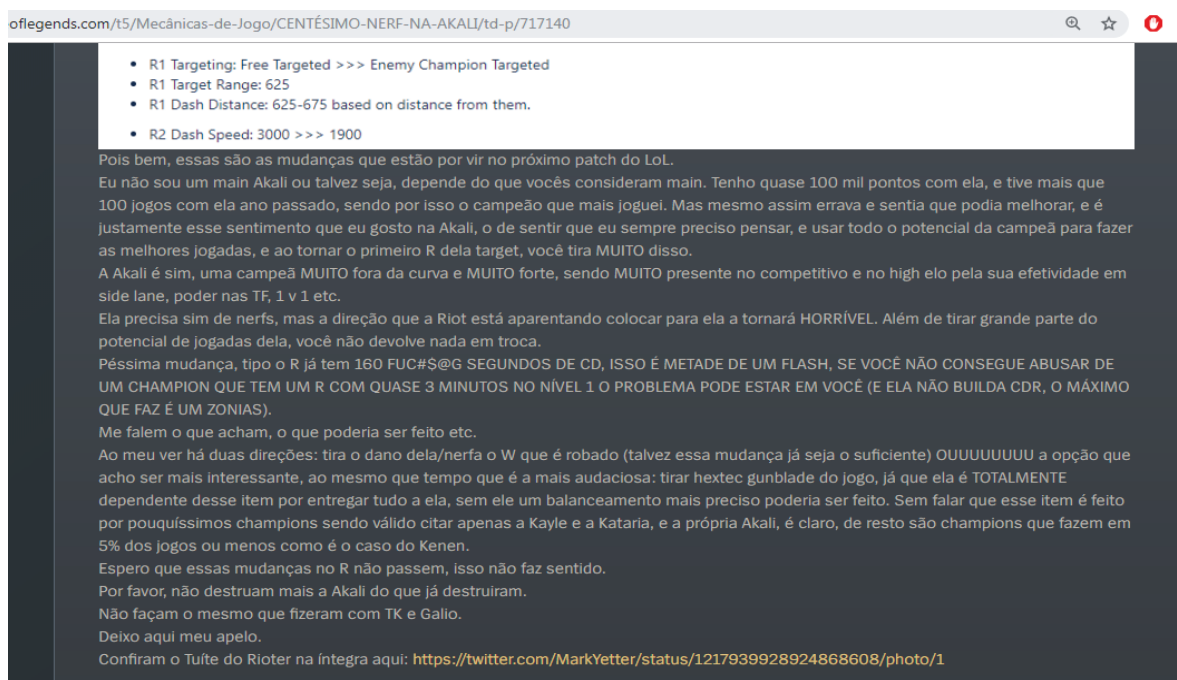
Figura 3: Seções de fóruns no website de *LoL*®



Fonte: Website *League of Legends*® (RIOT, 2020).

Os comentários realizados por meio dos fóruns são discutidos por diversos participantes, assim, apresentando e evidenciando o vocabulário utilizado. Dessa maneira, torna-se mais prática a coleta dos dados para a pesquisa. Na figura a seguir, observamos como são as postagens feitas pelos jogadores:

Figura 4: Postagem em fórum



Fonte: *Website League of Legends®* (RIOT, 2020).

1.4 O que se encontra em termos linguísticos nas interações dos fóruns

Ao estarmos em contato com a linguagem produzida por esse público, podemos refletir sobre como jovens brasileiros, que dificilmente possuem contato com aulas de língua inglesa, tampouco com falantes da língua, podem ter acesso e utilizar palavras que são de origem estrangeira dentro de suas interações.

Vemos que, do ponto de vista linguístico, fenômenos parecidos ocorrem em comunidades nas quais existem duas línguas em contato, as quais podem ou não possuir proximidades gramaticais e fonológicas, porém, tais características não se aplicam ao ambiente de interação que não é bilíngue, posto que os participantes dos fóruns são, geralmente, jovens brasileiros, de acordo com o levantamento realizado sobre o público que tem contato com jogos eletrônicos.

Assim, podemos pensar em uma relação de uso de palavras em uma língua estrangeira e suas misturas com a língua portuguesa, devido ao contexto de uso e motivado pelo *game* que possui seu vocabulário em língua inglesa, fomentando, dessa forma, o seu uso por falantes brasileiros que não são bilíngues nem possuem proficiência na língua estrangeira.

Sendo assim, a linguagem que esses jogadores empregam nesses ambientes é o foco desta pesquisa e, em consonância com a discussão acima, este trabalho pretende responder aos seguintes questionamentos:

1. Quais são os fenômenos relacionados às misturas de códigos linguísticos observados nas interações entre os jogadores nesse fórum *online*?
2. Quais são as recorrências de uso desses fenômenos e o que eles nos dizem sobre a linguagem no contexto da interação entre os jogadores e usuários do fórum *online*?

Para entendermos os fenômenos presentes nessas interações, compilamos um *corpus* e, para a investigação dos resultados, utilizamos as teorizações acerca de *code-mixing* (CM) (MUYSKEN; SCIULLO; SINGH, 1986; MUYSKEN, 2001). Para entender o que pode ser considerado como CM, precisamos também compreender as proposições acerca dos conceitos de empréstimo e *code-switching* (CS) (WEINREICH, 1968; POPLACK, 1980, 1982; POPLACK; SANKOFF, 2004; AUER, 1998; CACCAMO, 1998; GROSEJAN, 2006; GARDNER-CHLOROS, 2008; BAKER, 2001).

De acordo com os estudos postulados pelos autores mencionados acima, o CM e o CS existem em contextos em que há “fronteiras linguísticas”² ou comunidades linguísticas diferentes que estão em contato, porém, em um sentido físico e espacial.

Nesta pesquisa, no entanto, ponderamos que, apesar de não haver uma fronteira linguística material, o ambiente virtual atua como tal, de forma que a língua materna (L1) influencia a língua estrangeira (L2), assim como a L2 influencia na aquisição da L1.

É relevante, para a nossa área de atuação, a observação dos fenômenos intrínsecos à linguagem, dessa maneira, é importante salientar que esta pesquisa tem como ponto-chave a observação e coleta de dados sobre o fenômeno relacionado às fronteiras de linguagem, posto que, mesmo não sendo geograficamente ligadas, a fronteira digital nos permite estratégias peculiares aos dois idiomas, não sendo possível sua ocorrência em outros tipos de contexto.

² O termo “fronteira de linguagem” ou “fronteiras linguísticas” é abordado por Appel e Muysken (2005).

Tomando essa fronteira digital como alusão, Fiorin (2008) aponta que existem fenômenos curiosos de ortografia; no entanto, não é fácil encontrar palavras ou uma nova gramática na internet.

Fiorin (2008) tece um breve, mas importante, comentário sobre a linguagem, apontando que ela possui um fundo léxico comum, criado a partir de diversas misturas que foram sendo incorporadas à linguagem por culturas em contato, porém, o léxico comum não está sendo afetado por empréstimos, os quais, inclusive, são trazidos à língua portuguesa em sua primeira conjugação, que é a única conjugação produtiva no português moderno (FIORIN, 2008, pg. 3).

Para a organização empírica deste trabalho, definimos os objetivos gerais e específicos que permeiam o processo de pesquisa, tanto na coleta de dados quanto em sua análise:

Objetivos gerais:

- I. Observar o léxico empregado pelos jogadores de *LoL*® no período de março de 2018 a agosto de 2019, por meio do levantamento de *corpus*, tendo como foco os processos de uso de *CM* nos fóruns *online*.
- II. Analisar os dados coletados de acordo com o levantamento, procurando estruturas que sejam recorrentes ao uso das palavras que sofrem mistura de códigos linguísticos.

Objetivos específicos:

- I. Compilar o *corpus* de estudo que contém textos retirados de interações por meio de fóruns entre jogadores de *LoL*®;
- II. Analisar as interações linguísticas por escrito dos jogadores de *LoL*® nos fóruns *online*, procurando fenômenos linguísticos relacionados à mistura de códigos ou à sua alternância;
- III. Observar quais são as categorias de mistura de códigos mais empregadas pelos jogadores e qual a sua constituição entre L1 e L2;
- IV. Observar quais os agrupamentos ou *clusters* são recorrentes no *corpus* no que se refere aos *code-mixing* e empréstimos e quais são as ligações gramaticais presentes.

1.5 Organização do trabalho

Na presente seção, apresentamos a organização deste trabalho, pontuando de maneira breve o conteúdo de cada seção e suas subseções, em consonância com os propósitos desta pesquisa.

Na seção 1, está organizada a revisão bibliográfica deste trabalho, ou seja, apresentamos o arcabouço teórico que fundamenta a pesquisa. O fenômeno que envolve a mistura de códigos linguísticos é presente no ambiente de interação dos jogadores, sendo assim, é visto que, nos estudos sobre empréstimos, existe o que chamamos de *code-mixing*, ou seja, a junção e uso de dois códigos linguísticos pelo falante. A fim de entendermos a forma como os jogadores utilizam essas estruturas mescladas, foi necessário compilar um *corpus* que possuisse as produções linguísticas desses jogadores, dessa maneira, nos atemos aos pressupostos metodológicos de coleta e análise oferecidos pela LC.

Na seção 2, abordamos a forma como compilamos o material para o *corpus*, assim como mostramos a ferramenta que utilizamos para a compilação, a saber, o programa *AntConc*® (ANTHONY, 2019). Também expomos os critérios de seleção utilizados para compilar o *corpus* e como essa compilação foi realizada.

Na seção 3, apresentamos a verificação dos dados coletados, salientando a metodologia de análise. Inicialmente, exibimos a pesquisa realizada no *corpus*, contendo o número de palavras compiladas, a frequência de palavras e suas classificações, considerando as misturas de códigos linguísticos e os agrupamentos lexicais que são mais recorrentes. Ponderamos sobre os resultados obtidos, apontando, por meio do levantamento estatístico, os *CM* e empréstimos adaptados mais frequentes encontrados no *corpus*, observando suas ligações gramaticais.

Na última seção, podemos observar se os objetivos deste trabalho foram alcançados, discutindo-os e exibindo os passos futuros relacionados à pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, apresentamos as teorias que embasam esta pesquisa, as quais se concentram em duas áreas de estudos: o *Code-Switching/Code-mixing* e Empréstimos e os estudos da LC.

Nos estudos relacionados às línguas em contato, encontramos uma grande área de estudos associadas à mistura, alternância e “interferência” de códigos linguísticos, abordados teoricamente nos trabalhos sobre empréstimos (WEINREICH, 1968; PFAFF, 1980; POPLACK, 1982; HOFFER, 2002), *Code-switching* (CS) (PFAFF, 1979; POPLACK, 1982; POPLACK; SANKOFF, 2004; GARDNER-CHLOROS, 2009; WEI, 1998) e *Code-mixing* (CM) (POPLACK; SANKOFF, 2004; MUYSKEN, 2001; APPEL; MUYSKEN, 2005; UGOT, 2010), que pode ser observada na subseção 2.1 deste trabalho.

Discutimos as teorias oferecidas por esses estudiosos em consonância com exemplos observados no *corpus* compilado e, posteriormente, na seção de análise, retomamos alguns desses teóricos, pontuando os fenômenos que podem ser explicados por meio da teoria oferecida.

Por meio da LC (SINCLAIR, 1991; MCENERY; WILSON, 2001; BERBER SARDINHA, 2002, 2004), atemo-nos aos processos de compilação e análise postulados por essa linha teórica e apoiamo-nos no aparato teórico-metodológico oferecido para a seleção de dados, observação e análise dos resultados referentes ao estudo dos fenômenos de *code-mixing* e de empréstimos.

2.1 Os estudos das alternâncias e misturas de códigos linguísticos

A teoria que envolve o CM é uma corrente teórica nova em comparação aos estudos que já foram desenvolvidos em relação ao CS. Temos a perspectiva de “alternância” de códigos abordada inicialmente por Weinreich (1968), que nos apresenta, por meio de análise de determinadas comunidades bilíngues, o termo de “interferência”, no qual códigos linguísticos se misturam no discurso.

Weinreich (1968) postula que o termo interferência pode ser implicado no uso de padrões de uma língua, replicado em outra, e pode ser sintático, fonético ou envolver algumas áreas do léxico. Contudo, à luz do estruturalismo, Weireich (1968) desconsidera os eventos sociológicos que circundam os processos de interferência, demonstrando que, pelo viés

estrutural, essas interferências denotam a falta de proficiência dos interlocutores, assim como os erros o fazem no processo de aquisição de L2.

Apesar de os fenômenos relacionados a misturas de códigos linguísticos terem sido desconsiderados como uma área de estudos pelo estruturalismo, as pesquisas relacionadas às “interferências” começam a ser novamente exploradas nas décadas de 1970 e 1980, por meio da análise sistêmica do uso de CS como uma estratégia no discurso e um estilo de fala (BLOM; GUMPERZ, 1972; POPLACK, 1980; GROSJEAN, 2008; CACCAMO, 1998); o CS passa, então, a ser explicado como algo que está além de um processo randômico de mistura de códigos, ele possui regras.

Vemos que o termo “interferência”, postulado por Weinreich (1968), já não faz mais sentido à luz desta nova vertente teórica, a qual se enquadra nos princípios da Sociolinguística, o que, porém, não exclui o aparato investigativo postulado pelo estruturalismo. O que se verifica nesse momento é o fomento de novas investigações que darão vazão à busca empírica por exemplos reais desse fenômeno linguístico. (BLOM; GUMPERZ, 1972; POPLACK, 1982).

2.2 O Code-switching

Poplack (1980) aponta, em seus estudos iniciais, que o CS pode ocorrer de maneira consciente nas comunidades linguísticas, como em seu estudo sobre as relações de interferência entre o inglês e o espanhol em comunidades porto-riquenhas em Nova York. Neste estudo, o autor (1980) demonstra recorrências de uso e alternância nas orações, afirmando que existe sempre uma relação entre a língua matriz e a língua embutida e identificando determinadas regras linguísticas que validam o uso dentro de um contexto comunicativo, ou seja, Poplack (1980) expõe que os CS podem ser teórica e gramaticalmente explicados.

Poplack (1980, 1982) assume o viés explicativo da Sociopragmática (BLOM; GUMPERZ, 1972). Sendo assim, os CS possuem razões gramaticais e regras para serem realizados. O pesquisador (1980) postula as teorias de “equivalência condicionada”³ e de “condição dos morfemas livres”⁴.

A teoria da equivalência condicionada trata-se uma área investigativa que apresenta regras de aplicação de CS muito mais elaboradas, com alternâncias “intrassentenciais”

³ O termo original na língua inglesa é “*Equivalence Constraint*”, postulado por Poplack (1979).

⁴ O termo original em língua inglesa é “*The Free morpheme Constraint*”, conforme Poplack (1979).

(POPLACK, 1980), ou seja, os interlocutores alternam entre duas línguas no espaço oracional, respeitando as regras dos dois idiomas, o que demanda dos interlocutores uma proficiência muito maior nas L1 e L2.

A teoria da condição dos morfemas livres diz respeito à alternância de constituintes em qualquer parte do discurso que não sejam prefixos ou sufixos. Nessa condição, não se abarca a produção fonológica, pois os interlocutores tendem a pronunciar as alternâncias de maneira próxima à L1. Dentro dessa condição, podemos observar que Poplack (1980) discute sobre alternância emblemática⁵ e alternância de substantivos, muito comum em falantes que não possuem um nível mais elevado de fluência, assim, é a partir dessa condição que Muysken (2001) introduz o que passa a ser considerado *CM*.

2.3 O *Code-Mixing*

Muita confusão em relação aos termos *CM* e empréstimo emergem quando o assunto é a mistura de códigos linguísticos, pois não há um consenso entre os estudiosos em relação a como definir de fato essas alternâncias ou misturas. De acordo com Poplack (2004):

Apesar da identidade etimológica com a língua doadora, os empréstimos estabelecidos assumem os níveis morfológicos, sintáticos, e frequentemente, fonológicos da língua recipiente. Eles tendem a ser recorrentes no discurso do indivíduo e se espalhar entre a comunidade. [...]. Empréstimos ainda diferem do CS, pois nele não há envolvimento da morfologia, sintaxe ou fonologia da língua doadora. (POPLACK, 2004, p. 590).

Consequentemente, mesmo tendo verificado não haver consenso entre os estudiosos, neste trabalho, abordamos o tema de maneira mais incisiva, adotando o uso de *CM* discutido por Muysken (2001). Consideramos que esse fenômeno pode ser observado no uso das combinações de inglês e português pelos jogadores, de tal forma que ocorre a inserção de palavras de origem estrangeira dentro de um contexto totalmente em português, assim como a adaptação de empréstimos.

Pfaff (1979) aborda que os empréstimos podem ser morfolologicamente adaptados, demonstrando que verbos em língua inglesa tendem a sofrer adaptações para serem utilizados em outra língua, por exemplo no espanhol usado pelas comunidades chicanas. Assim, da mesma forma, podemos encontrar esse método de adaptação também na língua portuguesa.

⁵ Originalmente chamados pela autora de *tag-switch* (POPLACK, 1980, p. 589), esses tipos de alternância ocorrem quando se utilizam expressões idiomáticas de L2 no discurso que está sendo realizado em L1.

Não existe um consenso em relação ao que pode ser caracterizado como *CS*, *CM* ou empréstimo, mas há, no entanto, nuances do que cada conceito pode abarcar.

O *CS* é relativamente mais antigo nos estudos da área sobre o bilinguismo, os *CM*, por sua vez, configuram parte de uma abordagem mais recente na área. Enquanto Poplack (1979, 1980, 2004) trata o *CS* como um conjunto de regras que considera as alternâncias sentenciais como verdadeiras alternâncias de código, ainda que deixando as misturas de códigos ou a troca entre pequenos conjuntos lexicais para serem tratadas posteriormente, Muysken (2001) passa a observar que a produção de misturas por bilíngues ou indivíduos no processo de aprendizado ou aquisição de L2 possui a mesma relevância para a área.

A partir deste pressuposto, Muysken, Sciullo e Singh (1986) cunham o termo *CM* ao se referirem a “misturas” que ocorrem em processos discursivos dentro de uma oração, sem alterar o tópico discursivo, assim, também, abarcando os níveis mais básicos das regras gramaticais.

Muysken (2001) analisa essa categoria de ocorrência mais profundamente, oferecendo explicações sociopragmáticas e psicolinguísticas que podem explicar a motivação de uso, assim observando as regras que recorrem no uso de *CM* pelos falantes bilíngues, estipulando três estilos de *CM*: Inserção, Alternância e Lexicalização Congruente.

O estilo de *CM* relacionado à Inserção (MUYSKEN, 2001) encaixa-se ao uso do léxico de L2, posicionado em um processo discursivo em L1. Neste nível de *CM*, essa inserção pode ocorrer em qualquer parte do discurso, desde que não burle as regras de ambas as línguas, ou seja, é necessário que seja gramaticalmente coerente para os interlocutores.

Tende-se a se conceber alguns equívocos em relação ao processo de Inserção (MUYSKEN, 2001) e em relação aos “empréstimos”, que geralmente se limitam a um único item lexical que possui um amplo uso dentro de uma comunidade.

De acordo com o exemplo acima, a inserção funcionaria no discurso, com a L1 sendo produzida inicialmente pelo interlocutor bilíngue em espanhol, inserindo, depois, uma construção lexical em língua inglesa e logo voltando para o espanhol.

Muysken (2001) mostra que a inserção de L2 no discurso, ocupando a posição pós-verbal, é uma prática recorrente e está ligada aos princípios postulados por Poplack (1979) em relação à condição dos morfemas livres, nos quais a inserção pode ocorrer após qualquer constituinte do discurso, desde que não viole as regras das duas línguas, ou seja, as regras gramaticais de ambas as línguas precisam ser respeitadas.

Em relação ao nível de Alternância proposto por Muysken (2001), este se relaciona à definição de Poplack (1980), que sugere alternâncias intersentenciais, ou seja, entre as frases.

No entanto, a Alternância, para Muysken (2001), se revela como sendo intrassentencial, ocorrendo nas frases produzidas pelos falantes, porém o autor aponta que, para esse nível de CM, é necessário um nível de proficiência maior.

Geralmente, espera-se que, nesta categoria de CM, o falante alterne diversos constituintes encadeados no discurso para outra língua, sendo este tipo de concepção o mais estudado entre os linguistas. Os constituintes alternados podem ser orações inteiras, partículas do discurso e advérbios (MUYSKEN, 2001, p. 97).

Já no terceiro nível, que compõe a Lexicalização Congruente, é o estilo de CM que compartilha a estrutura gramatical da sentença em partes ou completamente (MUYSKEN, 2001, p.122). Nesse nível, não passamos pela inserção ou alternância de constituintes, mas sim pela alternância e uso de sistemas gramaticais de L2 em L1, ou L1 em L2. Muysken (2001) aponta que existem diversos gatilhos que podem desencadear esse estilo de CM e apresenta dois principais: a abundância de palavras homófonas que ativam os CM e uma estrutura gramatical equivalente.

Vemos que a Lexicalização Congruente necessita que os interlocutores bilíngues possuam um nível maior de proficiência em ambas as línguas para que possam realizar essa mistura, o que é geralmente observado em comunidades bilíngues em que ambas as línguas - a língua matriz e a língua embutida - possuam prestígio, não havendo tradicionalmente uma separação entre elas (MUYSKEN, 2001, p. 9).

Nos próximos segmentos, discutimos em mais detalhes o que cada um desses estilos de CM representa, assim como apresentamos alguns exemplos retirados do *corpus* compilado para este trabalho.

2.3.1 Inserção

O processo de Inserção apresentado por Muysken (2001) e abordado acima é um processo que, segundo o autor, age como unificador em relação aos fenômenos relacionados aos empréstimos e “*nonce-borrowings*” observados por Poplack (1988) que se referem aos empréstimos que se integram à língua que os recebem. Muysken (2001) evidencia que existem Inserções de diversos tipos, apontando-nos quatro tipos que podem ser aplicados ao nosso estudo:

1. **Inserção de N⁶**: Apenas substantivos são inseridos;
2. **Inserção de NP⁷**: Predicados nominais são inseridos;
3. **Inserção de DGNP⁸**: Predicados nominais marcados por número, gênero e pronomes definidos são inseridos;
4. **Inserção de DP⁹**: Frases com determinantes completos.

Para analisar as características do *CM* por inserção, Muysken (2001) lança mão do Princípio da Adjacência, discutindo que, para que se possa aplicar este princípio, há a necessidade de se estabelecer a “constituência”, ou seja, observar os constituintes sintáticos como, por exemplo, itens lexicais ou orações preposicionais.

Vemos que, no *CM* insercional, são inseridos constituintes de outras línguas no discurso, o que é plausível apenas com palavras únicas. Todavia, se houver a inserção de mais de um constituinte no discurso, o *CM* não passa mais a ser por inserção, mas sim por alternância ou lexicalização congruente (MUYSKEN, 2001, p. 62). Abaixo, podemos ver um exemplo de Inserção dentro de um discurso retirado do *corpus* compilado para esta pesquisa:

*[...]sua função na tf é eliminar a **back lane**, seu time faz a **front lane** e você chega atingindo a **back lane**, adc/apc...*

No exemplo acima, temos a L1 sofrendo inserção por itens que são embutidos de uma L2, neste caso o inglês. Pode-se notar que o que é inserido, neste caso, são os adjuntos adnominais.

A maioria dos casos apresenta inserção de apenas um único constituinte do discurso. Muysken (2001) demonstra que existem, ainda, quatro características que devem ser consideradas neste estilo de *CM*, a saber:

Constituintes únicos: o *CM* por inserção é caracterizado quando há apenas a inserção de um constituinte de uma língua em outra língua. Por exemplo, Muysken (2001) afirma que as Inserções do espanhol em *Quechua* - um dialeto observado na Bolívia – são de constituintes únicos.

⁶ Lê-se N como *Noun*, cuja tradução é “Substantivo”.

⁷ Lê-se NP como *Noun Phrase*, cuja tradução é “Sintagmas Nominais”.

⁸ Segundo Muysken (2001), a sigla DGNP significa *Determiner Genre Number Phrase*, cuja tradução é “Sintagmas Determinantes de Gênero e Número”.

⁹ Entende-se DP como *Determiner Phrase*, os quais dizem respeito aos Sintagmas Determinantes de uma sentença.

Estrutura de aninhamento a b a: Neste caso, Muysken (2001) discute sobre a proximidade que as estruturas possuem gramaticalmente, assim, o constituinte que é inserido, deve possuir relação estrutural e gramatical com a língua que irá sofrer a inserção.

Itens lexicais: As inserções são em sua grande maioria feitas utilizando-se palavras com conteúdo lexical, as quais são muito mais presentes nesse tipo de *CM* do que palavras agramaticais, ou seja, sem peso lexical, como preposições, por exemplo.

Elementos selecionados: Vemos que os *CM* por inserção são geralmente mais objetos e complementos do que adjuntos.

Para chegar às características que o *CM* por inserção precisa seguir, Muysken (2001) analisou a língua Quechua e as inserções que eram realizadas em espanhol. De certa maneira, podemos perceber certa proximidade com o inglês quando inserido no português durante as interações entre os jogadores.

2.3.2 Alternância

Resumidamente, o *CM* por Alternância¹⁰ deriva do que Poplack (1988) define como *CS* intersentencial. Nele, as línguas são utilizadas em uma oração, mas se mantêm, de certa maneira, separadas, sem inserção de itens.

Muysken (2001) separa uma seção de seu livro para discutir o que seria, então, a Alternância em *CM* e observa que existem dois problemas relacionados à mistura de códigos: o primeiro deve-se às preposições e conjunções, e o segundo ao ponto de transição na alternância, quando a ordem das palavras nas duas línguas é paralela (MUYSKEN, 2001, p. 96).

O autor assume que a Alternância incorpora produções de L2 em L1, quando as palavras que serão inseridas não são palavras lexicais, assim como não se restringem a apenas um item, resolvendo o problema relacionado às preposições e advérbios, que são inseridos no discurso em relação ao ponto de transição, ou seja, onde as sentenças sofrem alternância.

Muysken (2001) aponta que a equivalência linear da ordem das palavras tem um papel importante no processamento da alternância, permitindo a justaposição das línguas, mesmo em diferentes partes da sentença (MUYSKEN, 2001, p. 114).

Assim como a Inserção, no processo de Alternância entre códigos linguísticos, devem-se seguir alguns requisitos, de acordo com Muysken (2001). O autor observa que na

¹⁰ A palavra Alternância será utilizada com a primeira letra em maiúsculo para se referir ao processo de *code-mixing* por Alternância, proposto por Muysken (2001).

Alternância é esperado que diversos constituintes sejam trocados dentro de uma oração, ou seja, podemos ter orações inteiras de L2 em um discurso em L1, como no exemplo: “Tell Larry que se calle la boca” (POPLACK, 1980, p.587)

Como segunda característica da Alternância, Muysken (2001, p. 97) aponta que as alternâncias não podem ser aninhadas, ou seja, não devem possuir relação gramatical e estrutural mútuas.

Outra característica observada em casos de alternância é a complexidade e duração das orações. Muysken (2001) aponta que, de certa maneira, quanto mais constituintes de outra língua são acionados no discurso, mais certos estaremos sobre haver um caso de *CM* por alternância.

Vemos que uma grande diversidade de elementos é intercalada com as palavras lexicais no processo de Alternância, como, por exemplo, preposições, advérbios, entre outros constituintes de uma língua estrangeira inseridos na língua matriz, os quais foram observados por Myers-Scotton (1992); assim como marcadores de discurso que não são sintaticamente integrados, ou seja, não possuem correlação com a língua na qual serão embutidos (MUYSKEN, 2001, p. 97-98).

2.3.3 Lexicalização Congruente

A Lexicalização Congruente aparece como um *CM* apontado por Van Hout e Muysken (1995) e é retomado e explicado com mais detalhes em Muysken (2001). Trata-se do terceiro *CM*, o tipo mais complexo e menos recorrente no processo de mistura de códigos.

Nesse processo, a mistura de códigos entre duas línguas leva-as a uma convergência, na qual os dialetos integram-se, como apontado inicialmente por Clyne (1972), cujo estudo observou imigrantes alemães e holandeses na Austrália e concluiu que similaridades entre as línguas facilitam o processo de *CM* e a posterior criação de dialetos.

Para que este tipo de *CM* ocorra, as línguas em contato precisam, em algum nível, gramatical e/ou lexical, compartilhar as mesmas estruturas. Para Muysken (2001), a lexicalização congruente tende a ser frequente quando duas línguas em contato apresentam abundância de palavras homófonas e que possuem estrutura sintática relacionadas, possibilitando o *CM* dentro do discurso, sem necessariamente haver correspondência lexical entre as línguas, como no exemplo a seguir entre o francês e holandês:

Bij mijn broer y a un ascenseur en alles.

At my brother's/there is an elevator/and everything.' (MUYSKEN, 2000, p.126)

Dentro da Lexicalização Congruente, alguns fenômenos podem ser observados em línguas diferentes, como, por exemplo, o emprego de empréstimos, criando componentes hibridizados no nível sentencial; no entanto, existem alguns elementos que categorizam de maneira mais adequada este tipo de *CM*, os quais são apresentados por Muysken (2001).

- **Equivalência estrutural e linear:** Nesta característica da Lexicalização Congruente, Muysken (2001) aponta que, como as variedades linguísticas são próximas, existe uma linearidade estrutural entre elas que permite que elementos sejam reorganizados nas sentenças, ocupando posições que são aceitáveis entre as duas línguas.

- **Code-mixing de multiconstituintes:** Quando se observa casos de *CM* por lexicalização congruente, espera-se que os itens possam ser inseridos em qualquer parte das orações, respeitando-se os limites sintáticos e estruturais compartilhados pelas duas línguas, ou seja, pode-se observar que diversos constituintes podem ser alternados, sendo essas palavras lexicais ou não (MUYSKEN, 2001, p. 129).

- **Mistura de não-constituintes:** Esta característica está relacionada com a anterior, ao mesmo tempo em que os constituintes podem ser alternados dentro de todo o discurso. Existe também a possibilidade de serem incluídas palavras que não são de conteúdo (palavras lexicais), como preposições, por exemplo.

- **Presença de estruturas não-aninhadas:** Assim como no tipo de *CM* por Alternância, o uso de estruturas não-aninhadas é comum na Lexicalização Congruente, ou seja, os elementos podem ser utilizados sem que haja nenhuma relação direta entre elas, como aponta Muysken (2001, p.130), não havendo necessidade de conexão gramatical entre o dialeto e a variedade padrão.

Qualquer constituinte pode ser alternado dentro da Lexicalização Congruente, assim a temos como o nível mais complexo e mais avançado de *CM* dentro de uma comunidade, geralmente apresentando-se em contextos bilíngues em que ambas as línguas ou dialetos possuam a mesma condição sociolinguística, sendo parecidos em sua estrutura e considerados de prestígio.

Os três tipos de *CM* apresentados acima estão relacionados às relações linguísticas das comunidades em que há o encontro de duas ou mais línguas, mas, no contexto apresentado

nesta pesquisa, não existem duas ou mais comunidades com línguas diferentes, mas sim o ambiente de interação e o jogo, que oferecem aos jogadores insumos necessários para a mistura de códigos e a sua manipulação.

Nos processos de empréstimos adaptados, obtém-se a transformação de substantivos em língua inglesa para verbos que são utilizados como “pertencentes” à língua portuguesa, como nos exemplos a seguir:

Farm (fazenda) – *farm*-ar
Ward (ala ou corredor) – *Ward*-ar

Basilio (2004) aponta que os processos de formação de palavras na língua portuguesa suportam esta regra, pois o mesmo processo exemplificado acima relaciona-se com processos naturais nos quais um substantivo transforma-se em verbo ou ao contrário, como exemplificado em Basilio (2004, pg 28):

“(1) a. João pôs o carimbo no atestado.
 b. João carimbou o atestado”. (BASILIO, 2004, pg. 28)

Temos um caso clássico de alteração de classe gramatical, que é substancialmente conhecido dentro dos estudos linguísticos relacionados à formação de palavras e de variações linguísticas. Basilio (2004) nos chama a atenção para o processo de sufixação observado na língua, no qual uma estrutura base recebe o afixo (sufixos ou prefixos).

A estrutura da forma derivada é a estrutura geral da adição de um afixo a uma base ou radical; a base é determinada gramatical ou semanticamente pelo afixo. Na sufixação temos a estrutura [[base] sufixo]x, em que o sufixo determina a categoria lexical X da palavra resultante. (BASILIO, 2004, pg. 23).

Nos processos de empréstimos adaptados, a sufixação da estrutura base para a alteração de classe gramatical é presente. Pode haver duas motivações intrínsecas a essas alterações, como aponta Basilio (2004): a primeira refere-se à necessidade de usar uma determinada palavra de uma classe gramatical específica em outra estrutura gramatical, e a segunda motivação se reflete na necessidade de aproveitar conceitos ocorrentes em uma classe gramatical em outra.

Na subseção a seguir, apontamos as teorias que se referem aos empréstimos adaptados e sua classificação dentro dos estudos relacionados à linguística.

2.4 Empréstimos adaptados

Outro aspecto importante que emerge dentre as misturas de códigos são os empréstimos, os quais geralmente são categorizados de maneira diferente de *CM* ou *CS*, e referem-se a unidades lexicais que já foram incorporadas à língua. Hoffer (2002) aponta que o termo “empréstimo” ou *borrowing* é predecessor do *CS* e *CM*, e diz respeito à tentativa de reproduzir em uma língua algum termo anteriormente encontrado em outra (HAUGEN apud HOFFER, 2002, p. 5).

Blom e Gumperz (1972), assim como Poplack (1982), passam a discorrer sobre os empréstimos como possuidores de características distintas de uso, a depender do nível sociolinguístico da comunidade que os utiliza, e, em seguida, sobre as teorias da linguística e da sociolinguística que os compuseram, aprofundando, assim, os estudos dos *CS*.

Os empréstimos são caracterizados por suas alterações linguísticas, geralmente fonológicas e morfológicas, diferenciando-se, portanto, dos *CM* e *CS* que mantêm suas características morfológicas e fonológicas, assim como semânticas.

Como observado acima, os empréstimos podem sofrer alterações linguísticas para que se incorporem à língua destino, por exemplo, a palavra *mix* (mistura) pode ser adaptada para a língua portuguesa como *mixar*, sofrendo a sufixação -ar, muito comumente utilizada para verbos de ação na língua portuguesa.

Pfaff (1979) aponta a existência do que podemos chamar de “empréstimos adaptados”, que condizem com a realidade sociolinguística dos indivíduos, apresentando, assim, regras de usos para estes fenômenos.

Pfaff (1979) discorre, ainda, sobre a diferença entre empréstimos adaptados e *CS*. Os empréstimos podem acontecer na fala daqueles que possuem competência monolíngue, enquanto o *CS* implica algum grau de competência bilíngue. (PFAFF, 1979, p. 295-296). Nesse sentido, tomamos essa asserção como guia para esta pesquisa, visto que o ambiente em que os jogadores brasileiros estão inseridos pode ser caracterizado como ambientes monolíngues de interação.

A seguir, apresentamos os pressupostos da LC necessários para que haja conformidade na coleta de dados e em sua análise, assim, podemos demonstrar detalhadamente quais são as práticas metodológicas adotadas nesta pesquisa.

2.5 Pressupostos da Linguística de *Corpus*

Para a análise de dados e sua compilação, seguimos os preceitos científicos de coleta de dados linguísticos, tendo a LC (SINCLAIR, 1991; BERBER SARDINHA, 2002, 2004; MCENERY; WILSON, 2001; BIBER; CONRAD e REPPEN; 1998) como nossa base teórico-metodológica.

Ao observarmos esses processos, podemos encontrar quais palavras são mais utilizadas, quais são as formas que sofrem misturas de códigos linguísticos e adaptações em empréstimos, assim como a quais outras palavras elas se ligam. A partir da obra de Berber Sardinha (2002, 2004), podemos ter uma noção mais precisa do que se trata um *corpus*, ele aponta, retomando Sanchez (1995) que:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SANCHEZ, 1995 p. 8-9, apud BERBER SARDINHA, 2000, p. 338).

Dessa maneira, nos organizamos de acordo com alguns preceitos importantes que devem ser seguidos. Abaixo, o quadro 1 apresenta informações sobre o *corpus* levantado nesta pesquisa, considerando a metodologia utilizada.

Quadro 1: Informações sobre o corpus

Corpus de fóruns <i>online</i> em português (LoL®)	Quantidade de palavras	Modalidade	Fonte
Total de <i>palavras em língua portuguesa</i>	30442	Escrito	Fóruns <i>online</i>

Fonte: Elaborado pelo autor

No quadro anterior, vemos o número de palavras – *tokens* (RIBEIRO, 2004; BERBER SARDINHA, 2004) coletadas em um *corpus*. Observa-se que, de acordo com a LC, quanto maior o *corpus*, mais representativo ele pode ser (SINCLAIR, 1991), porém, atualmente, a coleta de textos para a constituição de *corpus* estão mais voltados aos objetivos que se pretende atingir com a sua coleta.

Berber Sardinha (2004) aponta quatro pré-requisitos que devem pertencer a um *corpus* computadorizado, são eles: a *Autenticidade do Corpus*, a *Naturalidade*, os *Crêterios* e a *Representatividade*.

A Autenticidade do *Corpus*, segundo Berber Sardinha (2004, p. 19) representa os textos produzidos por falantes nativos que estão sendo compilados em um *corpus*. A Naturalidade, por sua vez, se relaciona à produção autêntica de linguagem, ou seja, os textos não podem ter sido criados artificialmente, portanto, a fonte dos textos deve vir de produções autênticas e naturais (BERBER SARDINHA, 2004, p. 19).

Um *corpus* também necessita de critérios pré-estabelecidos para que sua coleta compreenda as intenções do pesquisador e do objetivo de coleta deste *corpus* (BERBER SARDINHA, 2004, p.19). No caso desta pesquisa, os critérios de coleta¹¹ nos auxiliam a encontrar o máximo possível de produções de CM.

O *corpus* precisa ser representativo, ou seja, precisa ser construído para que represente as condições de uso de uma língua, porém, Berber Sardinha (2004) afirma que a representatividade pode ser caracterizada de diferentes formas e geralmente está ligada à extensão do *corpus* e à probabilidade de ocorrências de determinadas palavras.

Apesar de o *corpus* ter como uma de suas características a representatividade, geralmente ligada a um grande número de palavras, a especificidade do *corpus* pode ser adequada aos tipos de pesquisa (BERBER SARDINHA, 2004).

O autor alega, também, que *corpora* pequenos tendem a ser mais representativos do que os subcorpora de grandes *corpora* (BERBER SARDINHA, 2000, p. 28). Sendo assim, nesta pesquisa, optamos por um *corpus* específico que contempla a linguagem escrita de jogadores de LoL®; especificamente, procuramos observar a utilização de palavras de língua inglesa dentro do discurso realizado em língua portuguesa.

O *corpus* compilado nesta pesquisa pode abranger a representatividade de um grupo específico de falantes, neste caso o de jogadores de LoL®. Berber Sardinha (2004) demonstra um parâmetro para se classificar os tamanhos dos *corpora*:

Quadro 2: Tamanho dos *corpora* e sua classificação

Tamanho em palavras	Classificação
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-grande
10 milhões ou mais	Grande

Fonte: BERBER SARDINHA, 2004, p. 26.

¹¹ Os critérios de coleta dos dados serão apresentados na seção Material e Métodos, no subtópico 3.2.

Discute-se que os *corpora* podem ter tamanhos variáveis de acordo com a sua necessidade e propósitos, como *corpora* para fins específicos, *corpora* de aprendizes, *corpora* pequenos, sendo este último o tipo no qual o *corpus* compilado neste presente trabalho enquadra-se.

2.5.1 *Corpora* pequenos

Os *corpora* pequenos, ou *small corpora*, são uma tendência crescente dentro da LC, distanciando-se da ideia de que o *corpus* necessita ser grande e possuir representatividade da língua na totalidade, conforme aponta Sinclair (1991, p. 18), ao afirmar que o *corpus* deve ser “o quão grande possível e, se possível, deve crescer com o tempo”¹².

Essa ideia de que para se representar melhor uma língua é necessário que haja o maior número de dados, e, no caso da LC, o maior número de palavras possível, perpetua-se dentro da LC desde o princípio.

Atualmente, a concepção de que quanto maior o número de palavras dentro de um *corpus*, mais preciso e melhor ele será, é difundida e defendida por pesquisadores da área. No entanto, essa concepção vem sendo debatida há algum tempo, como podemos observar nos estudos de Ghadessy, Henry e Roseberry (2001), em que discutem que os *corpora* pequenos podem ser utilizados com diferentes propósitos de acordo com os temas do estudo, ou seja, ensino, tradução etc.

Os *corpora* grandes permitem-nos observar diversos tipos de fenômenos que ocorrem na língua, pois abrangem um número grande de palavras, assim, sendo representativos de uma língua (SINCLAIR, 1991; MCENERY; WILSON, 2001), entretanto, *corpora* grandes podem não possuir termos específicos de uma língua, como variações de determinada palavra utilizada de maneira coloquial e que, raramente ou nunca, é transcrita nos textos que fazem parte do *corpus*, porém, torna-se possível acrescentar e analisar estas palavras por meio de *corpora* pequenos, nos quais podemos fazer recortes de determinadas comunidades linguísticas, falantes, grupos sociais etc. (GHADESSY; HENRY e ROSEBERRY, 2001).

¹² *The only guidance I would give is that a corpus should be as large as possible, and should keep on growing* (SINCLAIR, 1991, p. 18).

3 MATERIAL E MÉTODO

Nesta seção, discutimos os passos metodológicos adotados para coletar e analisar os dados desta pesquisa, a saber, comentários de fóruns do jogo *LoL*®. Para isso, dividimos esta seção em subseções, sendo a subseção 3.1 referente ao material utilizado na construção do *corpus*, ou seja, quais textos foram compilados.

Na subseção 3.2, apresentamos quais foram os critérios elaborados para a compilação do *corpus* e que nos guiaram em relação ao procedimento metodológico de coleta. A subseção 3.3 aborda quais foram os procedimentos de coleta do *corpus* e os passos metodológicos que adotamos para sua análise.

A subseção 3.4 contém as informações necessárias sobre o programa *AntConc*® utilizado como ferramenta para análise do *corpus*, na qual apresentamos as funções utilizadas para realizar a observação dos dados.

3.1 Material para construção do *corpus*

Como material de estudo para esta pesquisa, foram coletados textos de interações realizadas em um fórum *online* do jogo *LoL*®. Os textos foram retirados do *website forum.br.leagueoflegends.com*, o qual é atualizado diariamente pelos jogadores ou pela empresa criadora do jogo.

O *corpus* que foi coletado e analisado foi compilado no período de novembro de 2018 a março de 2019, ou seja, foram 5 cinco meses de coleta, o que nos ofereceu a quantidade de 30.442 palavras, incluindo palavras em português e inglês produzidas por falantes nativos de língua portuguesa.

A língua portuguesa foi o idioma utilizado pelos jogadores ao realizar os comentários nos fóruns. Levamos em consideração os aspectos relacionados à forma com que os jogadores se comunicam para a criação de critérios para a seleção desses textos, assim, pudemos ter acesso aos fenômenos linguísticos que esses indivíduos produziam em relação às alternâncias entre a língua portuguesa e a língua inglesa.

3.2 Critérios para coleta de material

O desenvolvimento metodológico desta pesquisa foi pautado nos critérios e procedimentos de coleta de dados e suas análises. Nesta seção, discutimos e demonstramos

quais serão os passos metodológicos que foram seguidos na compilação do *corpus* e em sua observação (SINCLAIR, 1991; BIEBER; CONRAD; REPPEN, 1998; BERBER SARDINHA, 2004; WEISSER, 2016).

O processo de compilação do *corpus* teve como objetivo observar o léxico empregado pelos jogadores de games *online*, nesse caso o jogo *LoL*®, observando se o *CM* se apresenta no ambiente de interação desses jogadores. Para a coleta destes dados, definimos os critérios de seleção de dados apresentados abaixo:

1. Fonte do texto
2. Conteúdo
3. Idioma do produtor do texto

O critério número 1 refere-se à fonte dos textos que são fóruns de discussão *online*, disponibilizados pela própria produtora do jogo. Nesse *website*, existe uma grande movimentação de usuários em postagens sobre problemas técnicos, dicas de como melhorar seus personagens, atualizações do jogo etc., dessa maneira, retiramos os textos de todas as seções de discussão dos jogadores.

O segundo critério se relaciona à quantidade de caracteres observados em cada texto coletado para que pudéssemos ter informações em quantidade satisfatória, tendo em vista que determinamos um número ideal de, no mínimo, duzentos caracteres por texto.

Inicialmente, consideramos esse critério impraticável, pois se mostrou demorado e pouco produtivo. Por isso, optamos por coletar todos os tipos de comentários, desde que fossem realizados pelos jogadores no fórum, excluindo textos pré-produzidos pelos administradores.

No terceiro e último critério, escolhemos textos produzidos por falantes da língua portuguesa, pois, por ser um *website*, podem existir postagens em outros idiomas, como inglês, por exemplo. Em consonância com nosso objetivo, pretendemos analisar a mistura de códigos linguísticos diferentes nesses ambientes de interação virtual, sendo vital que os produtores dos textos fossem falantes nativos de língua portuguesa.

3.3 Procedimentos de coleta do material

Os critérios de seleção dos dados devem considerar a autenticidade do texto. Por esse motivo, todos os textos coletados não sofreram alterações ou intervenção do pesquisador em

relação à forma escrita. Tais pressupostos oferecidos pela LC devem ser seguidos à risca para que os dados possam ser utilizados de maneira que sejam representativos de uma determinada comunidade linguística, nesse caso, os jogadores de *LoL*®.

Realizamos a coleta do *corpus* compilando os textos retirados dos fóruns (copiando e colando) e salvando-os em arquivos *Word*® no computador. Todos os textos foram refinados, retirando-se números ou elementos que não fossem necessários aos dados, como figuras e caracteres não linguísticos (*emoticons*, figuras e *gifs*). A partir deste ponto, convertemos todos os arquivos para o formato *txt*, para que fossem lidos no programa *AntConc*®, a fim de possibilitar a observação dos dados.

A partir dessa compilação das palavras e dos dados de frequência, separamo-nas de acordo com o seu uso (*CM* e empréstimos) e transferimos esses dados para o *Excel*®, assim, obtivemos a organização das palavras e das frequências de cada uma delas.

Nas próximas subseções, apresentamos quais foram os passos metodológicos para a avaliação dos dados coletados para este *corpus*. Na subseção 3.4, discorremos sobre a avaliação das palavras pela ferramenta *Wordlist* do programa *AntConc*®, por meio do qual averiguamos a existência de palavras de língua inglesa inseridas no discurso.

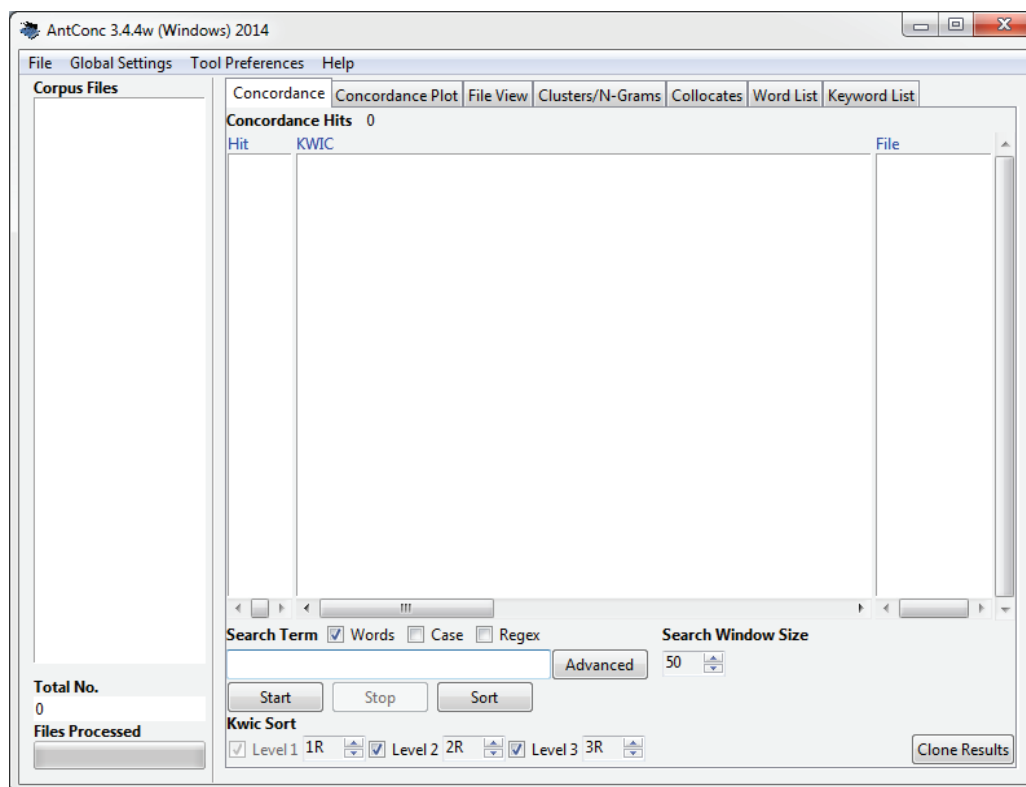
A seguir, apresentaremos o programa *AntConc*®, bem como suas funcionalidades e ferramentas para a compilação e estudo do *corpus*.

3.4 O programa *AntConc*®.

O programa *AntConc*® é uma ferramenta de uso gratuito e livre criada por Anthony (2019), utilizada para a compilação e análise de *corpora*. Nesse programa é possível utilizar as ferramentas *Wordlist*, *Concordance*, *Keywords List*, *Collocates*, *Concordance Plot* e *Clusters/N-gram*.

Na análise do *corpus* deste estudo, utilizamos a ferramenta *Word List*, *Concordance*, *Collocates* e *Clusters/N-grams* ao analisarmos a frequência, utilização e combinações de palavras nos textos.

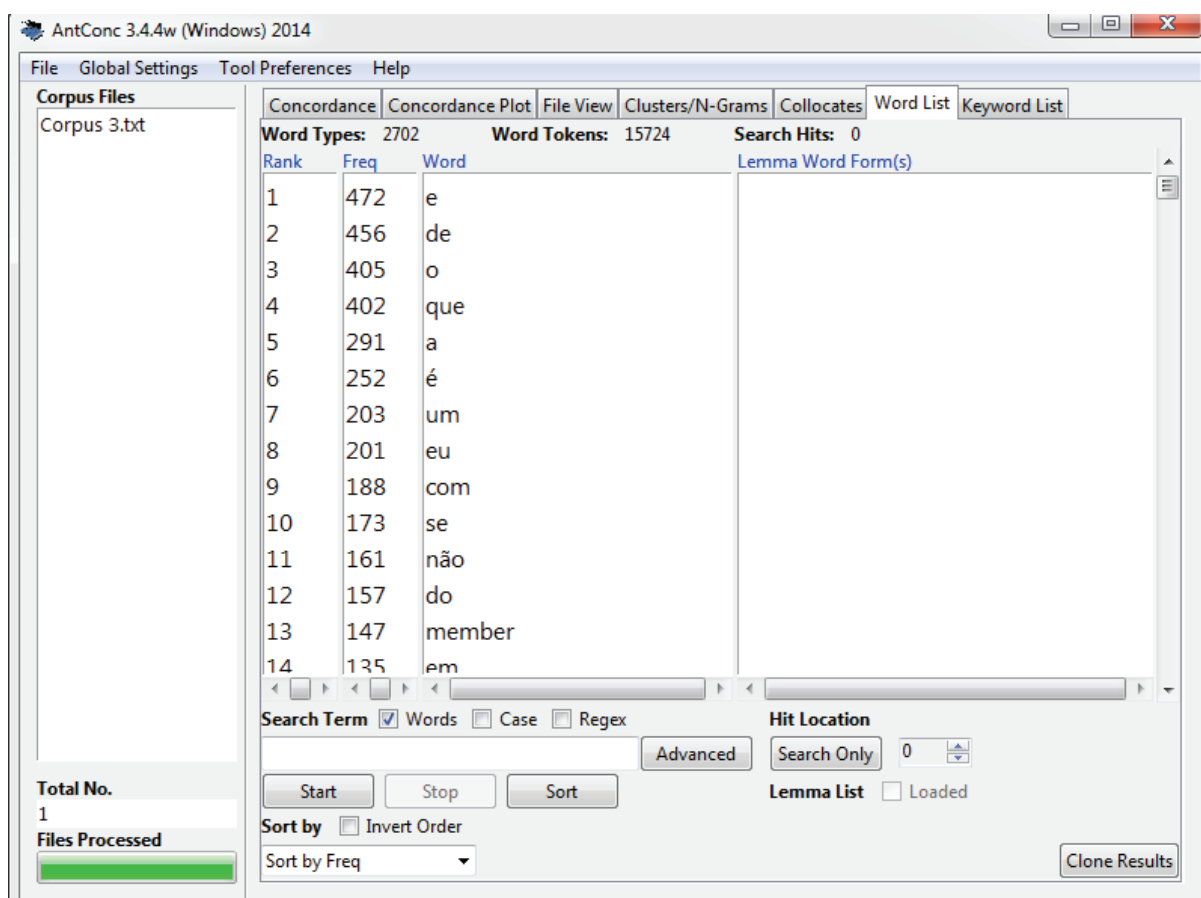
O programa possui uma interface amigável para os estudiosos da área de Linguística, não sendo necessário um conhecimento aprofundado sobre a LC para utilizá-lo. Como podemos ver na figura 5, a seguir, temos o *layout* inicial do programa *AntConc*®, com as opções que podemos utilizar para o *corpus* e as diversas ferramentas que podem auxiliar neste processo.

Figura 5: Layout inicial do programa *AntConc*®

Fonte: Captura de tela do programa *AntConc*®.

Na ferramenta *Wordlist*, pode-se analisar a frequência dos *tokens*, ou seja, das palavras que compõem o *corpus* e que se repetem diversas vezes ao longo dos textos coletados e inseridos no programa. Sendo assim, nessa ferramenta, são observadas as listas de palavras sozinhas, contendo a frequência com que essas palavras aparecem, bem como a sua posição no *ranking*.

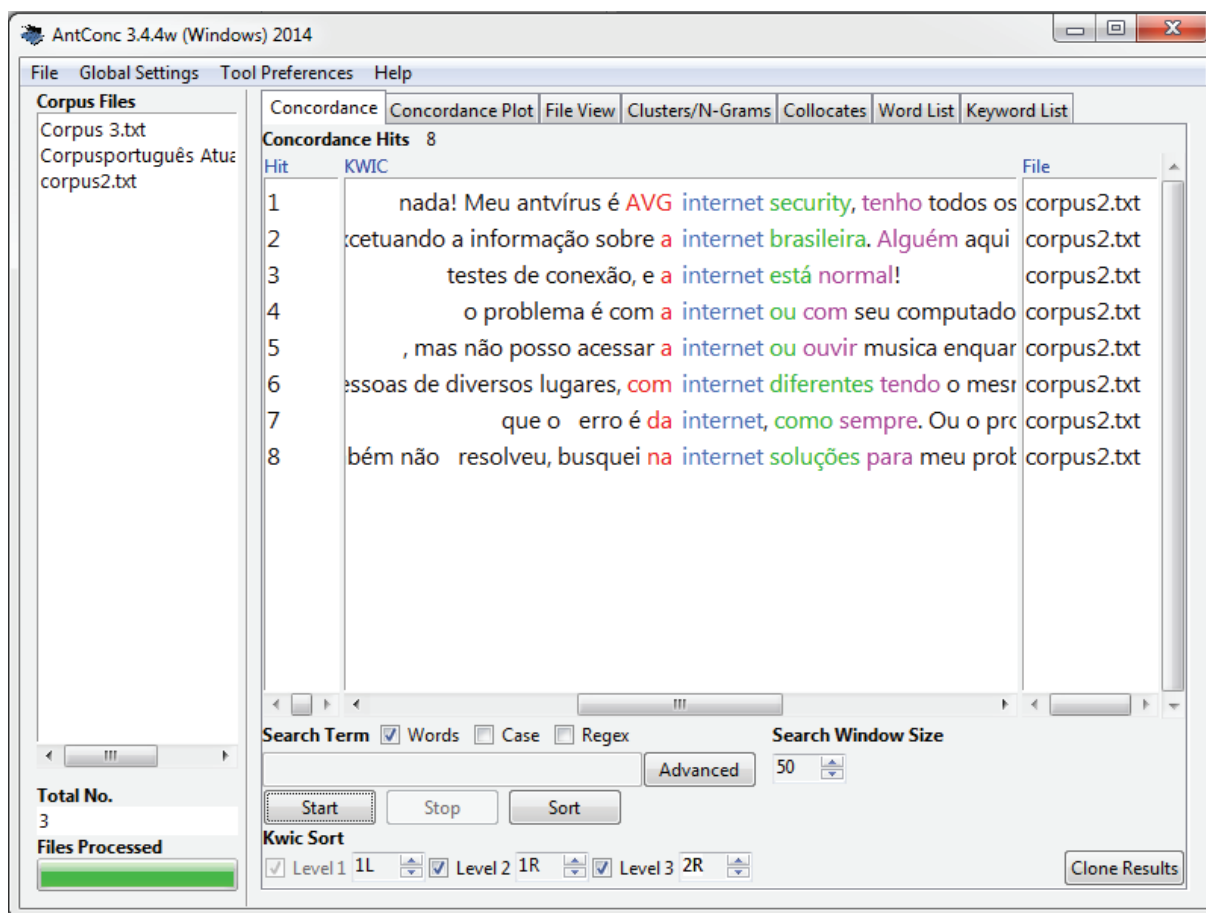
No programa, podemos organizar a lista de acordo com a necessidade do pesquisador, ou seja, se há o interesse em saber qual *token* é mais frequente, podemos deixar a lista de palavras organizada em ordem decrescente; ou, se o pesquisador quer saber sobre determinadas palavras organizadas por ordem alfabética, ele pode adequar a listagem dessa maneira. Outra forma, a qual foi adotada para esta pesquisa, consiste na ordem das palavras de acordo com a frequência de uso, como pode ser observado na figura 6, abaixo:

Figura 6: Ferramenta *Wordlist* do programa *AntConc*®

Fonte: Captura de tela do programa *AntConc*®

A ferramenta *Concordance*, ou concordanciador, é útil para que possamos entender qual a posição de uma determinada palavra dentro do texto ou em qual contexto a palavra insere-se. Por meio dessa ferramenta, é possível analisar quais agrupamentos de palavras podem ser observados. Também pode-se ver quais palavras estão em volta da palavra centralizada, assim como identificar e categorizar, por meio de cores, as funções gramaticais que as palavras exercem em relação ao contexto que está em volta do item “concordanciado”.

Logo abaixo, temos a palavra “internet” sendo centralizada pela ferramenta *Concordance*. Vemos que as linhas de concordância são alinhadas de forma que a palavra “internet” seja centralizada no programa. Observa-se que a palavra é “rodeada” pelas sentenças às quais pertence e vemos que o artigo “a” é muito recorrente junto a ela. Esse pode ser um exemplo de como a ferramenta *Concordance* pode ser útil ao linguista para encontrar padrões de uso na língua.

Figura 7: Ferramenta *Concordance* no programa *AntConc*®

Fonte: Captura de tela do programa *AntConc*®.

Na ferramenta *Collocates*, podemos ter as colocações de uma determinada palavra filtrando-as por frequências: frequência de tokens à direita ou à esquerda, e, também, pelo coeficiente estatístico, ou seja, a relevância que uma palavra tem em relação ao todo; no entanto, para os princípios desta pesquisa, nos ateremos à frequência que as colocações possuem.

Vemos, na figura 8, um exemplo de uso da ferramenta *Collocates* no programa *AntConc*®, no qual a palavra de busca continua sendo “internet”. Observa-se que, no topo da lista, está o artigo “a” com a maior frequência e, logo abaixo, constam os demais itens.

Figura 8: Ferramenta *Collocates* no programa *AntConc*®

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

- Corpus 3.txt
- Corpusportuguês Atualizado
- corpus2.txt

Total No. of Collocate Types: 22 Total No. of Collocate Tokens: 27

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	4	4	0	4.52909	a
2	2	0	2	4.81096	ou
3	2	1	1	4.32725	com
4	1	0	1	6.40192	tenho
5	1	0	1	8.55393	tendo
6	1	0	1	10.72385	soluções
7	1	0	1	5.38400	sempre
8	1	0	1	11.72385	security
9	1	0	1	3.74657	para
10	1	0	1	11.72385	ouvir
11	1	0	1	7.81696	normal
12	1	0	1	3.54394	no
13	1	1	0	3.99593	na
14	1	0	1	11.72385	hm

Search Term ☒ Words ☐ Case ☐ Regex

internet

Window Span ☐ Same From... 1L To... 2R

Min. Collocate Frequency 1

Start Stop Sort

Sort by ☐ Invert Order

Sort by Freq

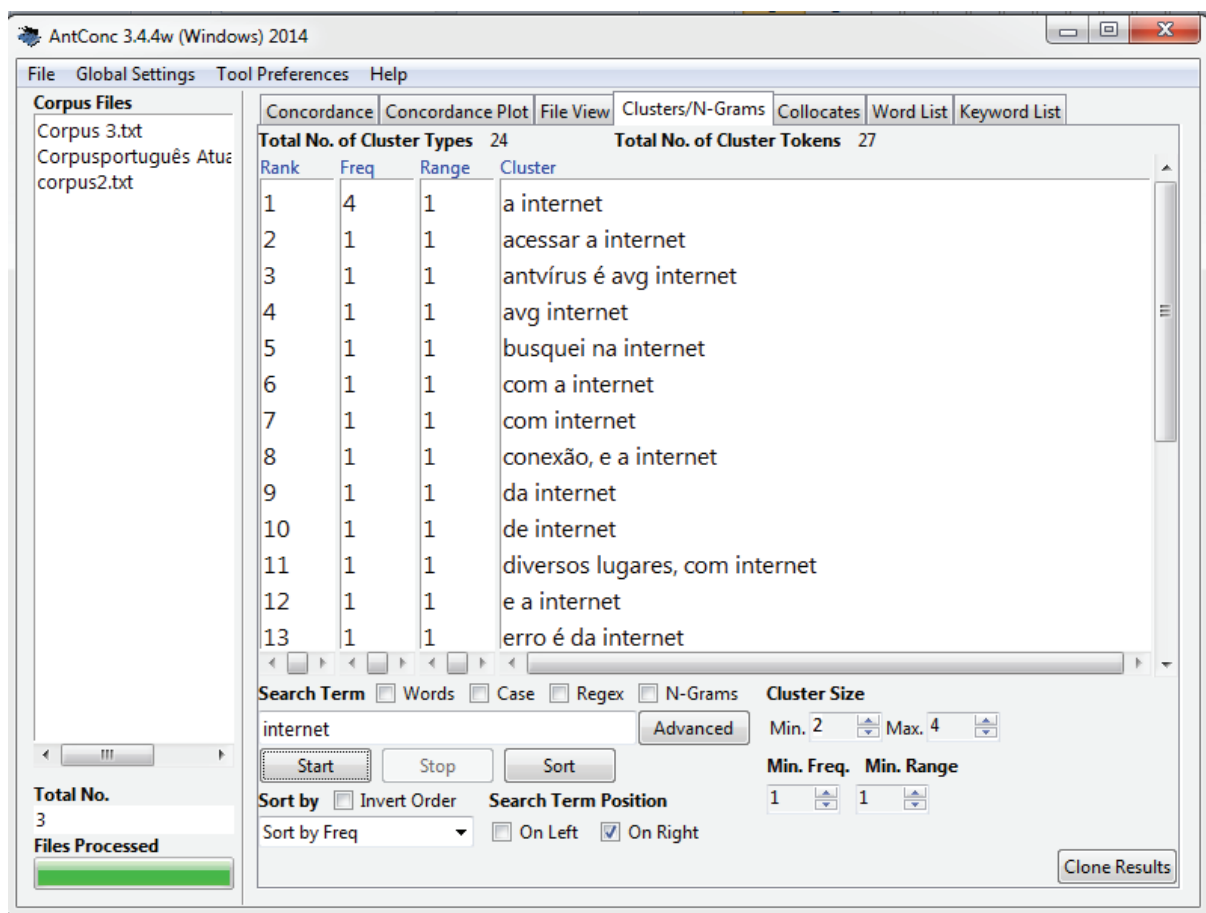
Clone Results

Fonte: Captura de tela do programa AntConc®

A ferramenta *Clusters/N-grams* é utilizada para saber quais são os “agrupamentos lexicais” de palavras, ou seja, quais as combinações que uma palavra pode possuir gramaticalmente, por exemplo: verbo + substantivo, verbo + advérbio, preposição + verbo ou advérbio etc.

Dentro desses aglomerados, temos os *N-grams* ou N-gramas, ou seja, as combinações de uma palavra com suas respectivas ligações. Temos, nessa ferramenta, uma maneira de identificar quais são as construções gramaticais mais presentes no *corpus*, podendo salientar quais são as classes de palavras que geralmente as acompanham e que são construções tipicamente frequentes na língua, como, por exemplo, “a internet caiu”. Veremos, a seguir, um exemplo retirado do *corpus* contendo a palavra *internet* como termo a ser buscado.

Figura 9: A ferramenta Clusters/N-grams no programa *AntConc*®

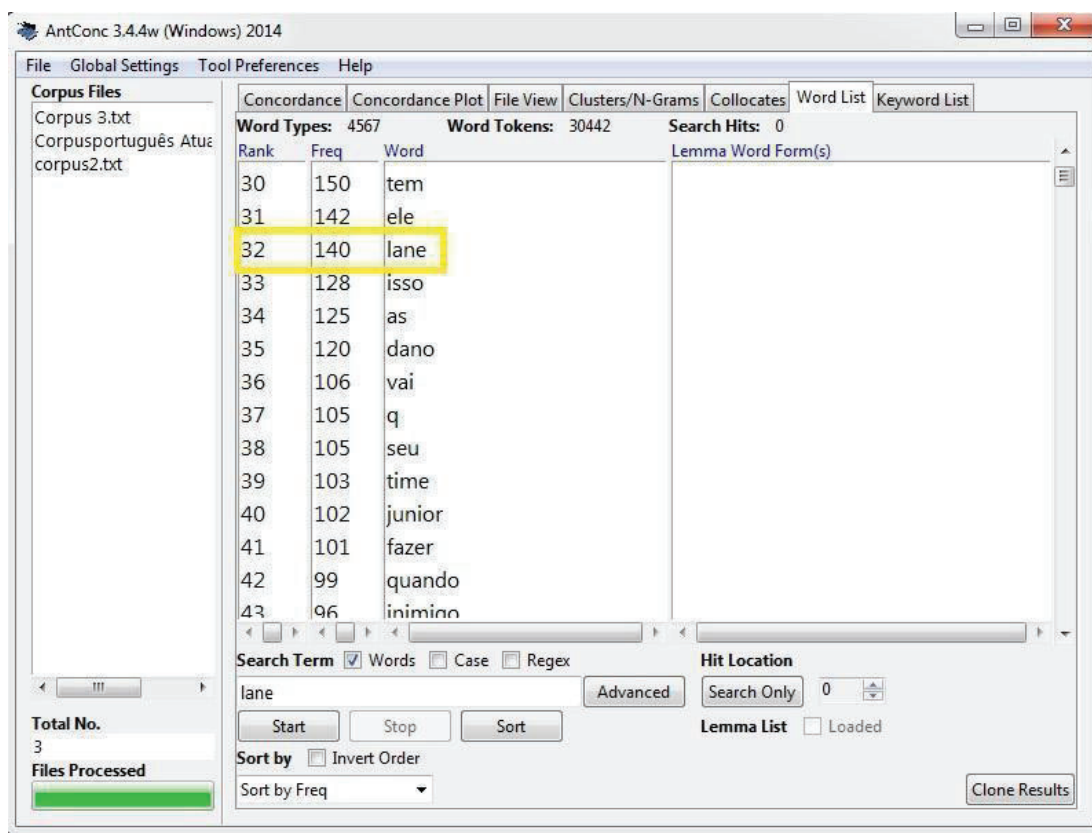


Fonte: Captura de tela do programa *AntConc*®

3.4.1 Avaliação das palavras pela ferramenta *Wordlist* (Lista de Palavras)

No primeiro momento, avaliamos a presença de itens agramaticais, como por exemplo, artigos e preposições.

A partir disso, passamos a checar os itens lexicais que procuramos, ou seja, palavras de origem inglesa que foram inseridas no discurso por meio de alternância, assim como palavras que têm o radical em língua inglesa, mas que, no entanto, sofreram processos de *CM*, misturando códigos linguísticos no interior da palavra e não na sentença.

Figura 10: *Wordlist* pelo programa *AntConc*®

Fonte: Captura de tela do programa *AntConc*®

Observamos na imagem anterior que, ao criarmos uma *wordlist* ou lista de palavras, podemos ter a frequência com que um determinado item se repete nos textos coletados. Vemos, na linha 32, que a palavra *lane* é repetida 140 vezes, o que nos indica uma frequência alta dentro de um *corpus* pequeno como o que compilamos.

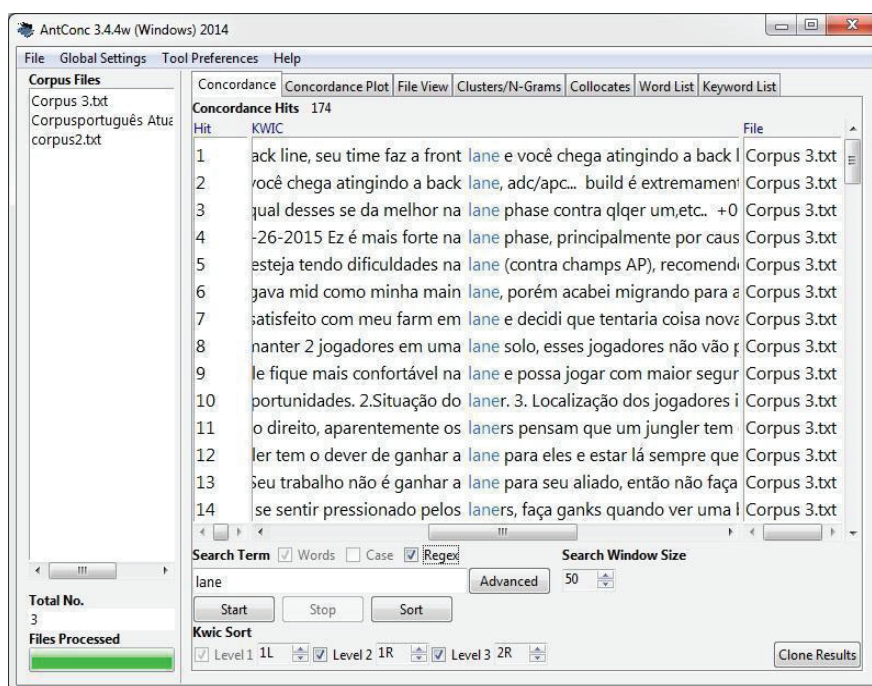
Tendo acesso ao item que nos interessa dentro da *wordlist*, analisamos qual das funções de *CM* essa palavra possui dentro do contexto, com base em Muysken (2001). Para isso, foi selecionado o item a ser analisado dentro das linhas de concordância.

A palavra *lane* foi selecionada porque está em língua inglesa no contexto em língua portuguesa. Segundo o *Longman Dictionary* (2009), *lane* significa, do inglês para o português, “pista, caminho, faixa”. Pode-se observar, no entanto, que essa palavra de língua inglesa foi inserida em um diálogo de língua portuguesa, o que poderia apontar um dos fenômenos estudados neste trabalho.

Ao utilizarmos a ferramenta *Concordance*, ou concordanciador, apresentada na figura 11 a seguir, torna-se evidente quais palavras rodeiam o termo central que retiramos da *wordlist*, indicando-nos, assim, o cotexto¹³.

O *token* centralizado é o que estamos analisando, assim, ele está destacado em azul e, ao seu redor, temos as sentenças às quais *lane* pertence, oferecendo-nos a oportunidade de analisar quais são as construções gramaticais ou lexicais em que essa palavra é inserida.

Figura 11: *Concordance* do programa *AntConc*®



Fonte: Captura de tela do programa *AntConc*®

Para ilustrar a frequência de uso dos empréstimos e dos *CM* no *corpus*, optamos por utilizar a lematização, ou seja, juntar as frequências de palavras para observar também quais eram as desinências dos itens levantados na compilação do *corpus*.

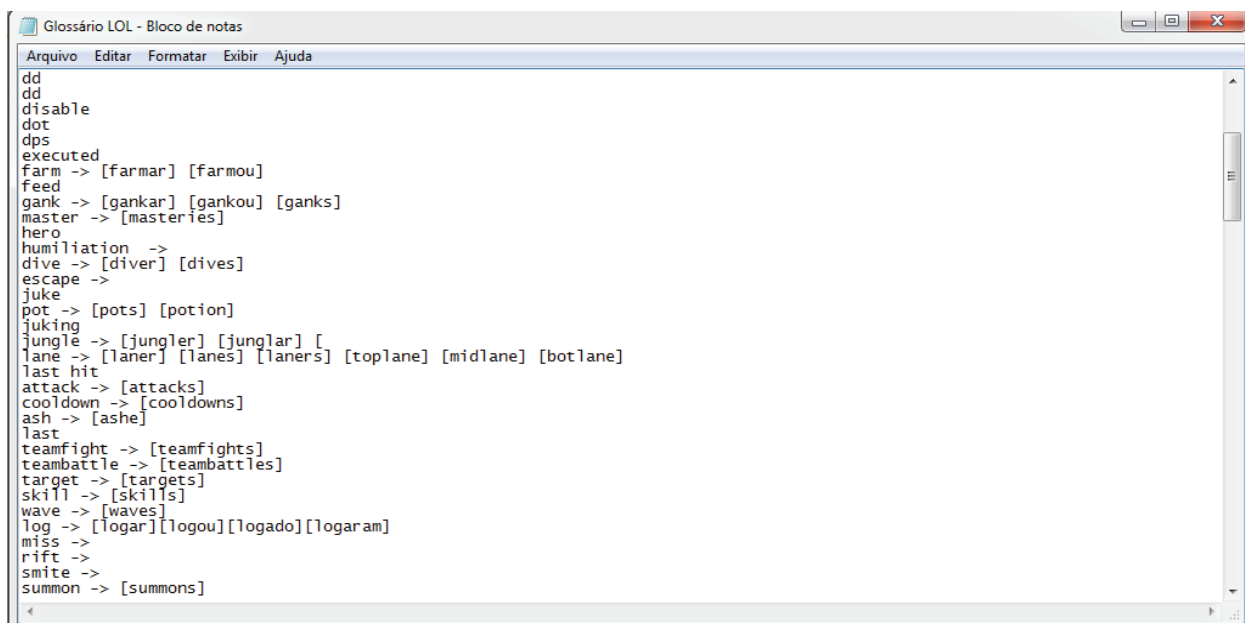
Em um segundo momento, foi criado um arquivo em *txt* que possuísse o vocábulo e suas variações gramaticais. Tais variações estão contidas no próprio *corpus* e são baseadas nas observações feitas em relação aos seus usos.

Com essa lista criada, utilizamos a ferramenta *Wordlist* para realizar o ajuste relacionado aos lemas que deveriam ser observados. Para isso, selecionamos o item *tool preference*; em seguida, optamos pela categoria *wordlist* que deu acesso à *lemma list*, em que

¹³ Cotexto para a Linguística de *Corpus* pode ser definido em relação às palavras subjacentes ao item centralizado em um concordanciador, dessa maneira, mostrando quais itens gramaticais e lexicais acompanham o termo.

adicionamos a nossa lista contendo as variações das palavras, como pode ser observado na figura 12.

Figura 12: Lista de variações (itens lematizados)



Fonte: Captura de tela de Bloco de Notas

A lista de palavras, com as suas variações, aparece ao lado direito da tela, apresentando também sua frequência no *corpus*. Na figura 13, a ferramenta *wordlist*, está apresentada com as variações das palavras à direita.

Figura 13: Lista de palavras (*wordlist*) no programa *AntConc*® com itens lematizados

corpus2.txt	47	88	bom	
	48	85	ser	
	49	83	pode	
	50	81	meu	
	51	81	sempre	
	52	79	gank	gank 44 gankar 12 gankou 1 ganks 22
	53	78	mesmo	
	54	77	jogar	
	55	76	adc	adc 76
	56	76	já	
	57	73	tempo	
	58	72	bem	
	59	71	melhor	
	60	70	ela	

Fonte: Captura de tela do programa *AntConc*®

Após esse passo, copiamos e colamos os dados no bloco de notas do *Windows(r)* com as palavras copiadas com o número de sua frequência, localizado logo à frente da palavra. Dessa forma, realizamos a formatação necessária para transferi-los para o programa Excel®.

Após a transferência, separamos o vocabulário em duas colunas, a primeira contendo empréstimos e a segunda contendo *CM*. Essas tabelas podem ser vistas na seção Anexos deste trabalho. Optamos pela função soma que pode oferecer-nos a soma de frequências absolutas dos *CM* e empréstimos compilados no *corpus*, possibilitando-nos passar para a fase de análise.

4 ANÁLISE E RESULTADO

Nesta seção, analisamos o material coletado e os resultados que obtivemos ao longo do período de pesquisa. Para melhor tratar desses pontos, a sessão foi subdividida em três partes, sendo a primeira, no item 4.1, a análise efetiva dos dados, informando o número de palavras coletadas, frequência de itens e os fenômenos linguísticos presentes.

No item 4.2, abordamos o que foi observado nos fenômenos de *CM* e de empréstimos presentes no *corpus*, explorando os agrupamentos lexicais relacionados aos *CM* e em relação aos empréstimos adaptados, assim como os conceitos teóricos que podem ser aplicados neste caso, oferecendo justificativas sobre o uso desta categoria linguística no ambiente virtual de interação.

O item 4.3 refere-se aos resultados obtidos durante a análise do *corpus* compilado, considerando o aparato teórico aqui discutido anteriormente.

4.1 A análise do corpus

A partir da observação da *wordlist*, procuramos todas as palavras que não pareciam ser vocábulos da língua portuguesa. Nesse passo, selecionamos as palavras e observamos as linhas de concordância nas quais se inseriram, quais eram os contextos e com quais palavras se relacionam. Assim, pudemos saber se as palavras possuíam radicais em língua inglesa e sufixação em língua portuguesa, como, por exemplo, em *gankar*. Sabemos que o radical da palavra vem da língua inglesa (*gank*), porém, foi adicionado a ela o sufixo -ar, tornando-a um verbo dentro do contexto comunicativo.

Após essa seleção de palavras, passamos a copiar e colar essas palavras dentro de planilhas no programa *Excel(r)*. Por meio desse procedimento, o programa ofereceu-nos a conta básica de quantas palavras tínhamos, tal como a porcentagem de frequência em relação ao *corpus* inteiro.

Após a realização desses processos, observamos que o uso de *CM* por inserção sobressaiu muito ao de “mistura” no nível morfofossintático, tendo, como amostra, a tabela ao final desta seção, com 863 ocorrências no *corpus* coletado, contra 409 ocorrências de empréstimos adaptados.

Para entendermos a frequência de *CM* e empréstimos, utilizamos os resultados oferecidos pelo programa *AntConc®*, os quais copiamos e colamos no Bloco de Notas, como pode ser observado na figura a seguir:

Figura 14: Frequência de empréstimos e *code-mixing* no corpus de transcrições de jogadores de MMORPG

Codemixing e empréstimos - Bloco de notas

Arquivo

Editar

Formatar

Exibir

Ajuda

botlane

1

lane

140

laner

23

laners

9

lanes

16

toplane

1

junglar

44

jungle

79

jungler

40

gank

64

gankar

12

gankou

1

ganks

22

game

33

farm

47

farmar

16

farmou

1

mid

41

middle

3

bot

40

top

24

buff

32

buffs

15

ult

33

ultar

7

early

11

skill

17

skills

20

ward

23

wardar

7

wardaram

1

wardou

1

kill

23

pentakill

1

champion

10

champions

12

ashe

15

tank

15

tanka

1

tankers

1

tanks

1

hit

15

hits

1

brush

15

carry

15

master

4

masteries

9

masterizei

1

poke

11

pokes

3

gold

13

leveis

1

level

12

burst

12

post

12

push

12

wave

4

waves

8

grave

11

graves

10

split

11

team

10

main

2

pot

2

pots

8

Fonte: Captura de tela do Bloco de Notas (Autor)

É possível verificar que, em cada linha, há uma palavra de entrada e, em seguida, o número referente à sua frequência, como no exemplo abaixo:

Lane > Botlane/laner/laners/lanes/toplane

Assim como foi observado na lista de palavras do programa *AntConc*®, o diferencial nesse arquivo foi sua distribuição, pois podemos observar que, em cada linha, há a palavra de entrada, assim como suas variações ao longo do *corpus*, e a frequência de cada uma delas.

4.1.1 Análise da frequência dos usos de *CM* e empréstimos

Temos, no programa *AntConc*®, o valor absoluto da frequência, ou seja, o valor de quantas vezes uma palavra apareceu no *corpus* coletado, o que pode ser visto na lista de palavras e na figura acima. Por exemplo, a palavra *kill* aparece 23 vezes no *corpus* coletado. Dessa forma, ela tem uma frequência absoluta de 23 vezes dentro das 30442 palavras, tamanho total do *corpus*, ou seja, cerca de 0,07%.

4.2 Frequência dos agrupamentos lexicais

Nesta subseção, discutimos os agrupamentos lexicais encontrados em relação aos *CM* e aos empréstimos e apresentamos suas recorrências. Para isso, dividimos esta subseção em subtópicos, relacionando a palavra e suas combinações.

A seguir, na seção 4.2.1, apresentamos as análises observadas com os *tokens lane, jungle, gank e farm*, nos quais os artigos e verbos ligam-se de forma majoritária aos empréstimos e *CM*.

Nas seções 4.2.2 e 4.2.3, apresentamos as investigações dos agrupamentos lexicais observados no *corpus*. Dividimos os dados da seguinte maneira: na seção 4.2.2, abordamos os agrupamentos lexicais envolvendo artigos + *CM* e, na seção 4.2.3, trazemos as locuções verbais presentes no *corpus*.

4.2.1 Agrupamentos lexicais (*clusters*)

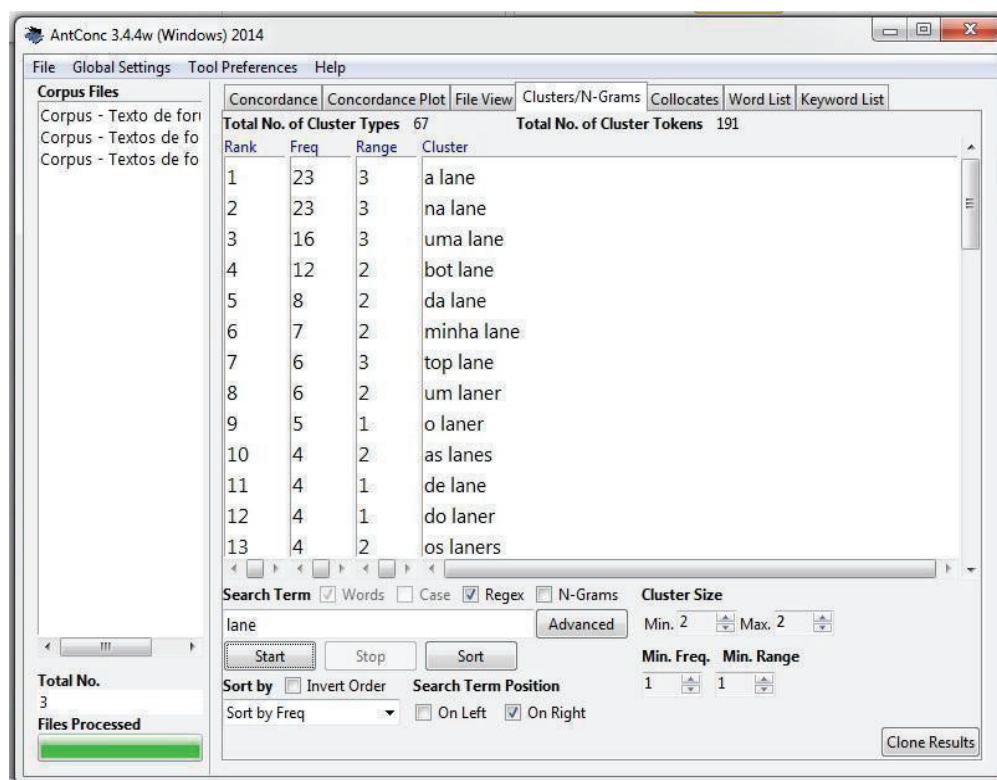
Nesta subseção, analisamos os *clusters* ou agrupamentos lexicais envolvendo os *tokens* mais frequentes observados no *corpus* compilado. Optamos por utilizar os *tokens lane, jungle, gank e farm*, para analisar suas ligações gramaticais devido a serem os mais frequentes e terem uma carga gramatical importante, caracterizando-se como substantivos e verbos. Os agrupamentos lexicais desses verbos e substantivos foram analisados levando em consideração as seguintes características:

1. Formas mais recorrentes que se ligam ao *token*;
2. Funções gramaticais que essas palavras desempenham;
3. Função gramatical que as palavras ocupam na mesma posição em um contexto em língua estrangeira.

Para analisar esses quatro *CM* mais frequentes, utilizamos a ferramenta *Clusters/N-grams* no programa *AntConc*®, abordada na seção de Materiais e Métodos. Essa ferramenta mostrou-se mais eficaz na forma de observar e analisar quais categorias de palavras e agrupamentos são mais recorrentes entre os *CM*.

A seguir, podemos ver um exemplo de uso para a ferramenta contendo o *CM lane*:

Figura 15: Ferramenta Clusters/N-grams

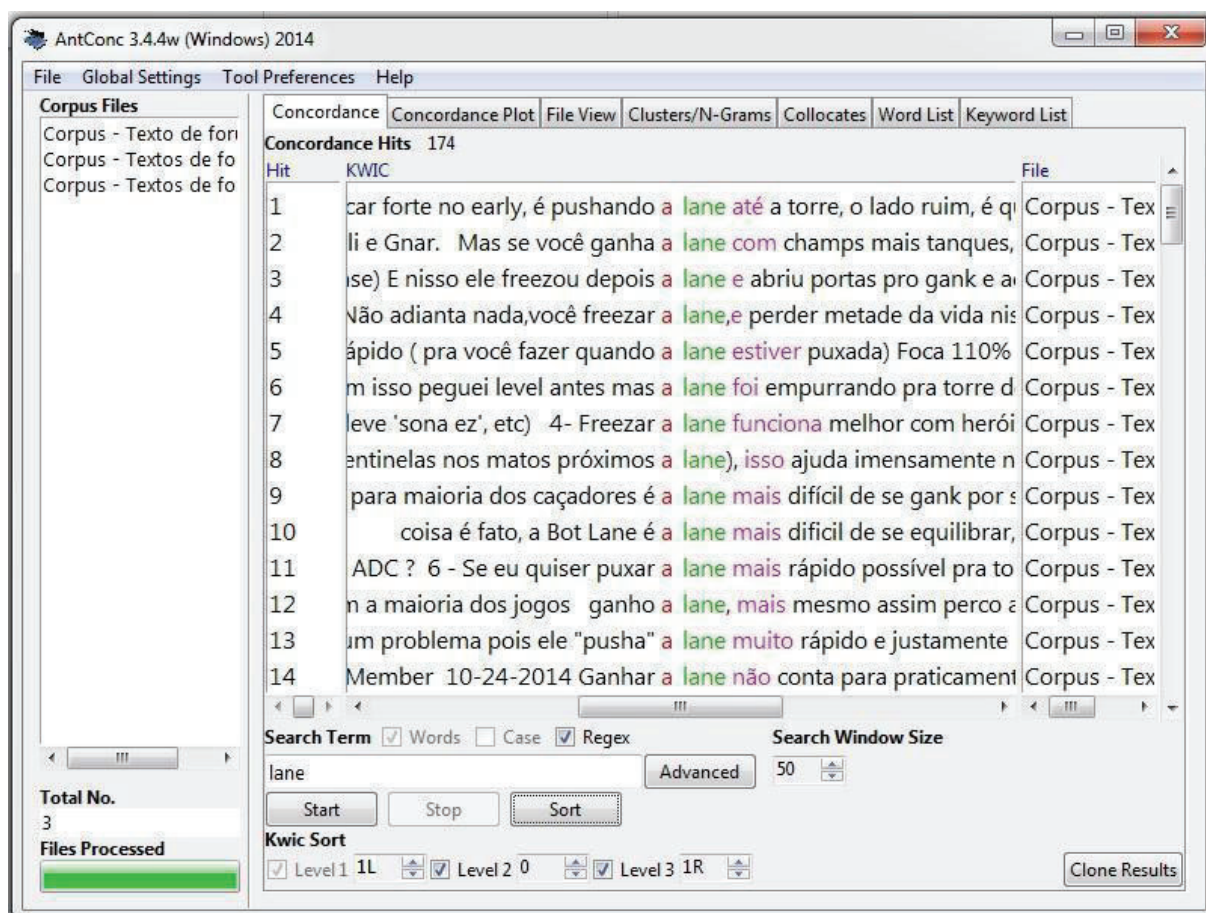


Fonte: Captura de tela do programa *AntConc*® (Autor)

Na figura 15, temos a palavra pesquisada *lane* em uma lista de frequência de combinações, vemos à esquerda a posição das palavras e sua frequência; à direita, observamos os agrupamentos lexicais. Foram essas as combinações de palavras e suas recorrências que selecionamos.

Outra ferramenta que nos possibilitou ter uma visão mais clara sobre o uso no contexto desses agrupamentos foi a *Concordance*, pois, ao centralizarmos as palavras que estamos buscando, podemos destacar e observar as palavras que vêm em volta, reconhecendo facilmente quais são suas funções.

Para analisar essa questão, observamos, primeiramente, a ferramenta *Concordance*, por meio da qual foi possível selecionar as palavras que acompanhavam o *CM*, baseando-nos nas cores destacadas, como o exemplo apresentado na figura a seguir:

Figura 16: Exemplo de combinações utilizando a ferramenta *Concordance*

Fonte: Captura de tela do programa *AntConc*® (Autor)

Pode-se notar que a palavra que está sendo pesquisada encontra-se centralizada em verde, as outras palavras no entorno possuem outras cores, como, por exemplo, o artigo definido “a” em vermelho e, depois da palavra centralizada, temos outras classificações de palavras destacadas em roxo.

Abaixo, apresentamos o quadro que informa os *tokens*, sua frequência e sua função gramatical na língua inglesa:

Quadro 3: *Tokens* mais frequentes de *code-mixing*

Token	Frequência	Função gramatical
Lane	140	Substantivo
Jungle	79	Substantivo
Gank	44	Verbo
Farm ¹⁴	33	Substantivo/verbo

Fonte: Elaborado pelo autor

¹⁴ *Farm*, de acordo com o *Longman Dictionary* (2009), possui a função gramatical de substantivo e verbo.

4.2.2 Artigo + *code-mixing* com função de substantivo

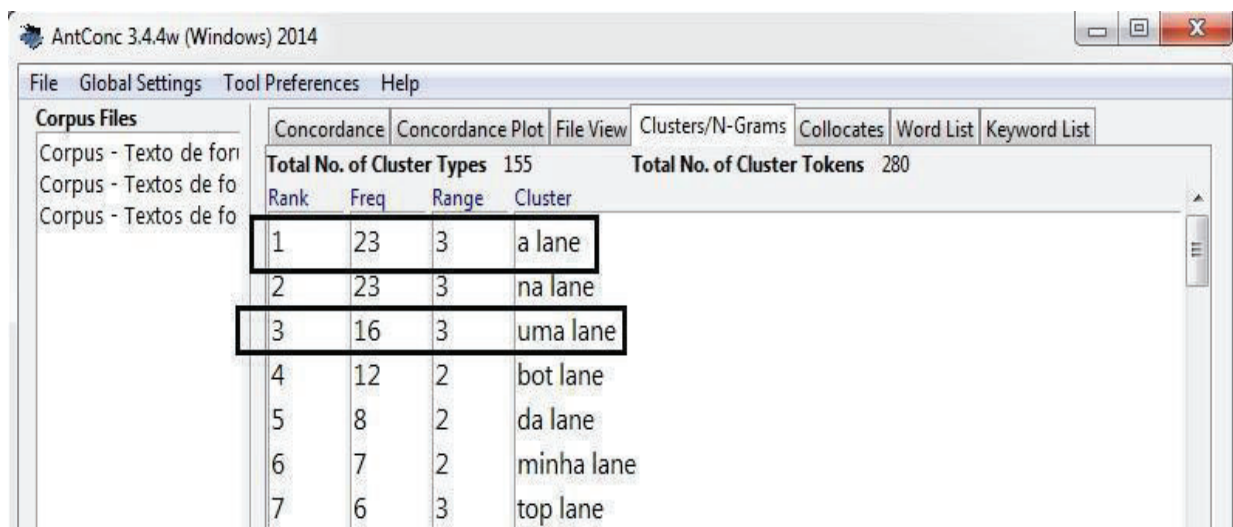
Ao realizarmos a análise dos agrupamentos, observamos que, inicialmente, as combinações lexicais mais recorrentes devem-se aos agrupamentos envolvendo artigos definidos e indefinidos¹⁵, como os *CM* apontados no quadro 3.

Vemos que *lane* e *jungle*, substantivos na língua inglesa, são os *CM* mais frequentes no *corpus* compilado, justamente por assumirem o papel de substantivos também em língua portuguesa.

Ao observar os agrupamentos dos dois substantivos, notamos que o uso de artigos é recorrente e frequente. Nota-se, ainda, que o artigo definido feminino “a” é o mais frequente, aparecendo 23 vezes com a palavra *lane* e 21 vezes com a palavra *jungle*, o que mostra que os jogadores reconhecem esses substantivos na língua portuguesa como sendo femininos.

Os artigos indefinidos são menos recorrentes nessa categoria de palavras, o artigo indefinido “uma”, por exemplo, tem a frequência de uso com o *CM lane* apenas 16 vezes. Nas duas figuras abaixo, podemos ver as frequências de uso desses artigos:

¹⁵ Bechara (2009) define artigos definidos como “simplesmente artigo o, a, os, as que se antepõem a substantivos, com reduzido valor semântico demonstrativo e com função precípua de adjunto desses substantivos.” (BECHARA, 2009, p. 130). Já com relação aos artigos indefinidos, segundo Bechara (2009), “a tradição gramatical tem aproximado este verdadeiro artigo de um, uns, uma, umas, chamados artigos indefinidos” (BECHARA, 2009, p. 130).

Figura 17: Artigos definidos + *code-mixing*


AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

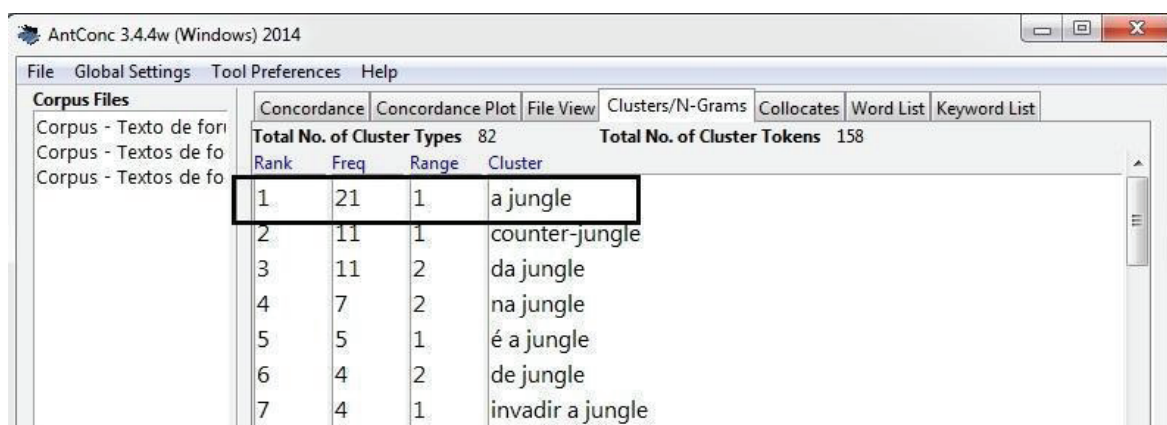
Corpus Files

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Cluster Types 155 Total No. of Cluster Tokens 280

Rank	Freq	Range	Cluster
1	23	3	a lane
2	23	3	na lane
3	16	3	uma lane
4	12	2	bot lane
5	8	2	da lane
6	7	2	minha lane
7	6	3	top lane

Fonte: Captura de tela do programa *AntConc*® (Autor)

Figura 18: Artigo indefinido + *code-mixing*


AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Cluster Types 82 Total No. of Cluster Tokens 158

Rank	Freq	Range	Cluster
1	21	1	a jungle
2	11	1	counter-jungle
3	11	2	da jungle
4	7	2	na jungle
5	5	1	é a jungle
6	4	2	de jungle
7	4	1	invadir a jungle

Fonte: Captura de tela do programa *AntConc*® (Autor)

No *corpus* compilado, podemos notar que esse padrão de frequência se repete com *CM* que são substantivos em sua língua de origem, ou seja, na língua inglesa. No quadro abaixo, os artigos definidos e indefinidos são apresentados nestes agrupamentos lexicais envolvendo *CM*:

Quadro 4: Frequência de uso de artigos definidos e indefinidos com *code-mixing*

<i>Code-mixing</i>	Artigo	Frequência de agrupamento
<i>Lane</i>	a	23
<i>Jungle</i>	a	23
<i>Kill</i>	a	7
<i>Burst</i>	um	6
<i>Skills</i>	as	6
<i>Hit</i>	o	4
<i>Carry</i>	o	4
<i>Gold</i>	ao	3
<i>Ward</i>	uma	3
<i>Champions</i>	os	2
<i>Skill</i>	uma	2
<i>Early</i>	o	3
<i>Gank</i>	o	3
<i>Farm</i>	a	1
<i>Poke</i>	o	1

Fonte: Elaborado pelo autor

No quadro, podemos observar que muitos desses *CM* possuem artigos definidos como formas recorrentes de combinação lexical, porém, outras formas são também usuais, como, por exemplo, os agrupamentos que envolvem os *CM* e empréstimos adaptados, tendo em seu agrupamento a utilização de verbos e locuções verbais. Bechara (2009) os define como:

Chama-se locução verbal a combinação das diversas formas de um verbo auxiliar com o infinitivo, gerúndio ou particípio de outro verbo que se chama principal: hei de estudar, estou estudando, tenho estudado. Muitas vezes o auxiliar empresta um matiz semântico ao verbo principal, dando origem aos chamados aspectos do verbo. (BECHARA, 2009, p. 191).

A partir dos exemplos de agrupamentos, vemos que os artigos são flexionados e ligados de maneira inusitada com as palavras que são inseridas no contexto comunicativo do jogador.

Tomemos como exemplo a palavra *jungle*, que, sendo um substantivo que não possui gênero na língua inglesa, traz para seu uso em contexto em L1 o gênero feminino. Dessa maneira, os jogadores optam por utilizar o artigo “a” para se referir a essa palavra, pois o uso criado por eles mesmos assim permite.

Por sua vez, a palavra *kill*, sendo um verbo e não possuindo gênero na língua inglesa, assume o papel de um substantivo feminino, pois é apenas utilizada para se referenciar à “morte” de um personagem. Sendo assim, o substantivo “morte” é designado como um substantivo de gênero feminino, o que pode explicar a opção de uso do artigo “a”, mesmo que não haja a tradução direta no jogo.

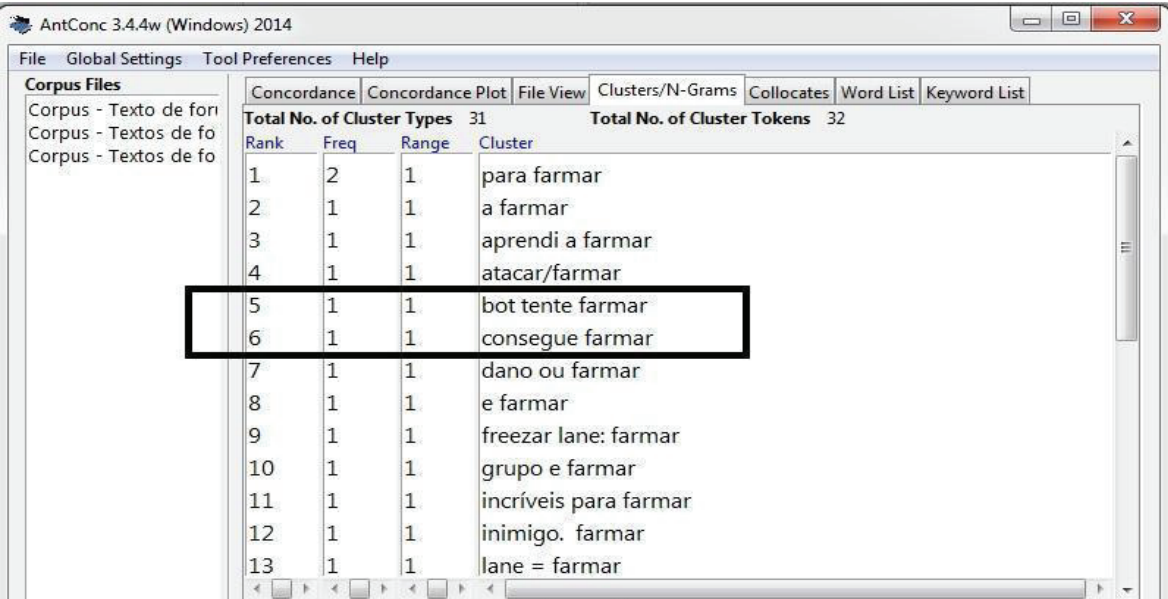
Na subseção a seguir, podemos ver exemplos de uso em que verbos são ligados aos *CM* e aos empréstimos.

4.2.3 Verbos + *code-mixing* e empréstimos

Discutimos anteriormente que o uso de artigos com *CM* na função de substantivo é recorrente dentro de nosso *corpus*, porém o fenômeno que mais se ressalta envolve os verbos em língua portuguesa e em língua inglesa. Vimos que os artigos definidos são, em sua maioria, ligados a substantivos em língua inglesa. Alguns verbos de língua inglesa podem assumir a posição de substantivo quando inseridos em um contexto de língua portuguesa, como, por exemplo, o *CM kill*.

O uso de verbos adaptados da língua inglesa para a língua portuguesa é recorrente, pois se observa no *corpus* que são mais frequentes do que o uso de substantivos. Outro fator que pudemos analisar é o uso de locuções verbais, quando verbos em língua portuguesa se unem a verbos de língua inglesa, os quais são verbos de ação, como pode ser observado na figura a seguir:

Figura 19: Locução verbal envolvendo empréstimo adaptado com sufixação -ar



Rank	Freq	Range	Cluster
1	2	1	para farmar
2	1	1	a farmar
3	1	1	aprendi a farmar
4	1	1	atacar/farmar
5	1	1	bot tente farmar
6	1	1	consegue farmar
7	1	1	dano ou farmar
8	1	1	e farmar
9	1	1	freezar lane: farmar
10	1	1	grupo e farmar
11	1	1	incríveis para farmar
12	1	1	inimigo. farmar
13	1	1	lane = farmar

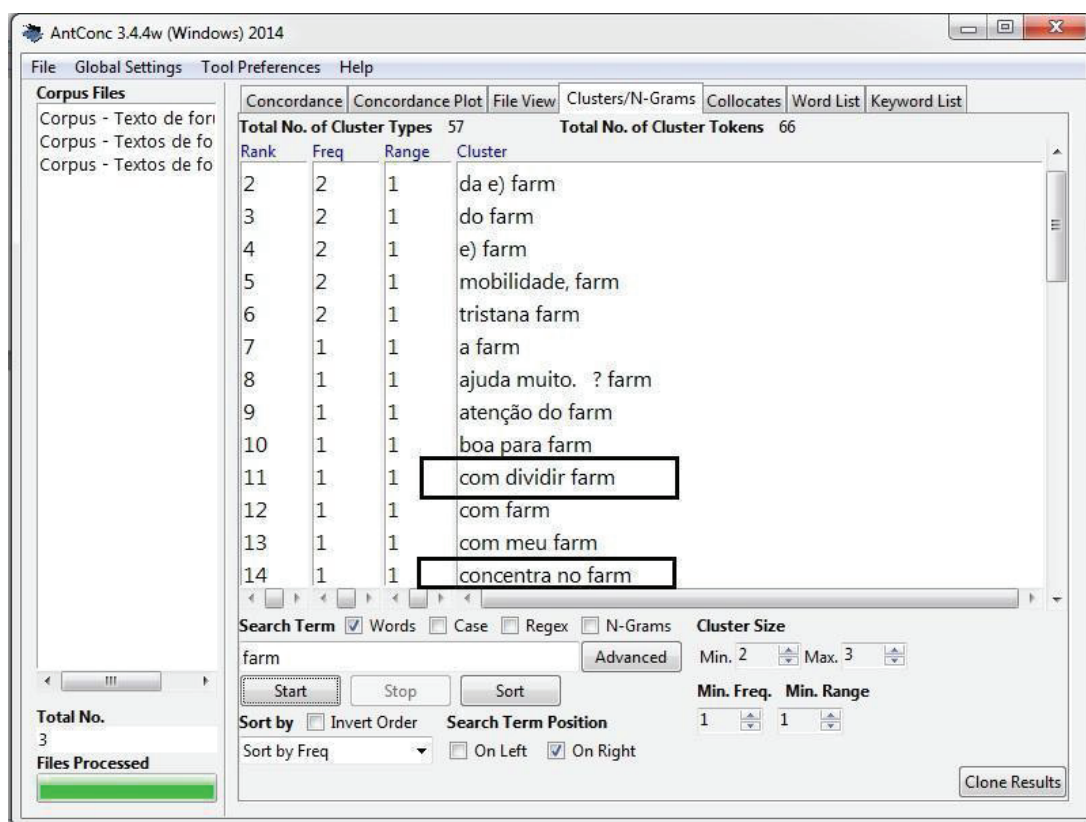
Fonte: Captura de tela do programa AntConc® (Autor)

Nesse caso, temos as locuções verbais *tente farmar* e *consegue farmar*, que estão presentes no vocabulário dos jogadores. A forma adaptada *farmar*, recebe a sufixação “-ar”, indicando que *farm* passa a ser um verbo de ação, o que é recorrente nas falas dos jogadores ao adaptar verbos e até mesmo substantivos de língua inglesa para a língua portuguesa.

Na seção Apêndice, podemos observar as duas tabelas contendo os *CM* e empréstimos adaptados. Pode-se perceber, no Apêndice B, que existem empréstimos que possuem a mesma construção, sempre recebendo a sufixação “-ar”.

Os casos em que ocorre locução verbal são mais voltados ao uso de adaptações morfológicas de empréstimos provindos da língua inglesa, porém elas não acontecem quando se trata de *CM* envolvendo os substantivos ou verbos, que são tratados geralmente como objeto direto ou indireto.

Figura 20: *CM* como objeto da ação



Fonte: Captura de tela do programa *AntConc*®

Para obtermos uma melhor noção sobre as ligações entre verbo e *CM* ou empréstimos, organizamos um quadro apresentando os verbos mais recorrentes entre os *CM* e empréstimos morfollogicamente adaptados. Neste quadro também apresentamos os empréstimos e *CM* agindo como objeto da ação do verbo, assim, a título de ilustração, organizamos o quadro utilizando as palavras mais recorrentes no *corpus*, dividindo-as com os verbos que as acompanham e que são mais recorrentes, nesse caso os verbos “tentar”, “fazer” e “dar”.

Quadro 5: Verbos recorrentes e suas ligações com os *CM* e empréstimos

Verbo	Code-mixing/ Empréstimo
Tentar	Farmar Gankar Wardar
Fazer	Ganks Counter Buffs Poke Peel Split
Dar	(last) Hit Counter Buff Counter Ganks Poke Bait Rage Split

Fonte: Elaborado pelo autor

No quadro, fica evidente que os três verbos mais recorrentes são verbos de ação, empregados pelos jogadores ao explicar suas jogadas e táticas no jogo. Esses verbos geralmente são sucedidos por preposições, como, por exemplo, “tentar *wardar as*” e “vou *gankar uma*” principalmente o verbo “dar”, mostrando que os *CM* e empréstimos adaptados assumem a função de objeto indireto. Isso se deve à própria língua portuguesa, na qual o verbo “dar” pode ter transitividade direta ou indireta, podendo ser sucedido por uma preposição, geralmente “a”.

Para ilustrar a frequência de uso dos empréstimos e dos *CM* no *corpus*, decidimos, inicialmente, fazer a separação manualmente. Para isso, observamos toda a *wordlist* à procura dos termos a fim de contabilizá-los. Os números, na tabela a seguir, mostram os itens contabilizados.

Ao realizar a contagem, percebemos o número mais elevado de palavras relacionadas às inserções de palavras de língua inglesa nos comentários postados pelos jogadores. Tal diferença pode ser observada no quadro 6, no qual apontamos a tipologia dos *tokens* (*CM* e empréstimos), assim como sua frequência.

Quadro 6: Quantidade e tipologia dos *tokens* observadas no corpus

<i>Empréstimos</i>	409	Escrito	Fóruns online
<i>Code-mixing por inserção</i>	863	Escrito	Fóruns online
Total de palavras coletadas.	30422	Escrito	Fóruns online

Fonte: Elaborado pelo autor

Como vimos anteriormente, os empréstimos adaptados são menos numerosos em relação aos *CM*, pois eles são substantivos e palavras que já existem no jogo, possuindo uma visibilidade e necessidade de uso maior do que as palavras que são adaptadas. Como exemplo, temos a palavra *jungle*, o mapa do jogo, a qual o jogador tende a usar por referir-se ao ambiente em que está jogando, e a palavra *gankar* (uma adaptação da palavra *gank*), que é utilizada apenas em jogadas ou comentários entre os jogadores.

Na seção a seguir, apresentaremos os resultados e discussão, onde mostramos que o uso dessas adaptações por parte dos jogadores é recorrente dentro de suas falas, o que poderá ser observado pelo diagrama de caixa, no qual algumas palavras são sobreutilizadas, enquanto outras são usadas de forma mínima.

4.3 Resultados e discussão

Podemos considerar que o fato de os empréstimos serem menos numerosos em relação aos *CM* se deve aos termos utilizados, como *CM* representarem habilidades, lugares e até mesmo ações no jogo que não podem ser traduzidas, a tendência é que seu uso seja recorrente. Um exemplo disso ocorre com a palavra *lane*, que foi utilizada 140 vezes no *corpus* compilado, ou seja, sua incidência é alta comparada com, por exemplo, sua forma hibridizada *laner*, que aparece 23 vezes.

Ao observar o contexto da palavra *lane*, veremos que ela ocupa estruturalmente o lugar de um substantivo, como observaremos no quadro 7, no comentário de um jogador postado no fórum sobre como realizar uma jogada utilizando um personagem do jogo.

Quadro 7: O uso de *code-mixing* e empréstimos no discurso dos jogadores

<i>Code-mixing</i> com função de substantivo.	“qual desses eu compro? Kog maw e late game?,ezreal e late game? qual desses se da melhor <i>na lane</i> phase contra qlqer um,etc..”
O empréstimo morfológicamente adaptado com função de substantivo.	“Maokai (que particularmente é meu tipo favorito), deixar a kill para o laner não é uma má ideia, se essa kill ajuda-lo a sobrepujar o oponente com mais facilidade ele pode se tornar menos dependente da sua ajuda.”

Fonte: Elaborado pelo autor

Observamos, no quadro acima, trechos retirados de comentários no fórum *online*. Observa-se que os itens em negrito são *CM* e empréstimos, o primeiro é a palavra *lane*, que possui correspondência com a língua inglesa e cujo significado, segundo o *Longman Dictionary* (2009), é “pista”; no entanto, o item “*laner*” não possui nenhum correspondente na língua inglesa, porém é utilizado frequentemente pelos jogadores para designar um personagem que geralmente inicia os ataques em uma *lane* do jogo.

Outro aspecto interessante que podemos observar em relação à compilação do *corpus* é em relação às categorias de *code-mixing* utilizadas. Muysken (2001) aponta, em seus estudos, a existência de três estilos de *CM*, a saber, inserção, alternância e lexicalização congruente, no entanto, no *corpus* pudemos observar apenas um estilo, o *CM* por inserção. Esse fenômeno pode ser explicado pela forma como a língua inglesa é explorada por esse público, por meio de palavras isoladas que se referem a itens comuns ao ambiente discursivo dos jogadores, o que pode explicar, também, a alta frequência dos *CM* em relação aos empréstimos.

Podemos considerar, então, que os empréstimos que incidem nessa comunidade discursiva são variações dos *CM*, passando inicialmente, como explica Poplack e Sankoff (2004), como “*nonce borrowing*”, ou seja, empréstimos que se integram à língua que os recebem.

Podemos observar novamente o caso do *CM lane*, que sofre adaptação para *laner*, ou, em casos mais complexos como o substantivo de origem inglesa *ward*, que recebe diversas variações tornando-se um verbo, sendo assimilado à língua portuguesa e recebendo desinências verbais próprias do português, como pode ser visto a seguir.

Vimos anteriormente que Pfaff (1979) discute os empréstimos que são adaptados morfológicamente quando se integram à língua de chegada. Nesse sentido, apresentamos a seguir a ocorrência de empréstimos, que é muito menor do que as de CM por inserção, como, por exemplo, em *ward*:

Quadro 8: Frequência de uso do *code-mixing ward*

Vocábulo	Frequência	Tipologia
<i>Ward</i>	17	Code-mixing por inserção
<i>Wardar</i>	7	Empréstimo
<i>wardaram</i>	1	Empréstimo
<i>wardou</i>	1	Empréstimo

Fonte: Elaborado pelo autor

Notamos ser discrepante o número de ocorrências entre a forma regular proveniente da língua inglesa em relação à forma hibridizada, morfológica e sintaticamente modificada do substantivo aplicado ao contexto discursivo dos jogadores, como pode ser visto no excerto retirado de uma das postagens no fórum que apresentaremos a seguir:

Quadro 9: Tipologia e exemplos

Tipologia	Excerto
Code-mixing	“Boa noite Meu nome é paulo igor e gosto de ficar no bot como suport. eu queria saber como colocar ward no local certo porque o povo reclama e fica pedindo "poe sentinela, onde é que ta a sentinela, tu num pos não" --' da raiva mais como não sei ao certo ponho perto dos matos só que o povo vem e retira não sei como faz isso mais eles fazem, se pude-se me ajuda eu ia ficar feliz ;PP obrigado”
Empréstimo	“O que fazer se você sofrer counter-jungle? Você não é perfeito, esta sujeito a sofrer com counter-jungle da mesma forma que seu oponente. Para lidar com counter-jungle existe um hack muito bom chamado; Sentinela de Visão/Ward tente wardar as entradas da sua selva, isso te da a possibilidade de ver quando algum filho da mãe esta entrando no seu mato.”

Fonte: Elaborado pelo autor

No primeiro excerto, que está categorizado como *CM*, vemos que o jogador usa a palavra *ward* como substantivo, por meio da construção “colocar *ward*”. Temos, aqui, o uso de *ward* na sua forma original, sem sofrer nenhuma alteração morfológica ou sintática, sendo assim, como explica Muysken (2001), um *CM*, porém, no segundo quadrado, categorizado

como empréstimo, o trecho retirado de uma postagem no fórum, há o uso da locução verbal “tente *wardar*”. Esse fenômeno trata-se de uma adaptação para a língua portuguesa do substantivo *ward* de radical na língua inglesa (*ward*), que sofre a sufixação “-ar”, transformando-se, assim, em um verbo de ação em língua portuguesa.

Existem diferenças entre os usos de *CM* e empréstimos no contexto dos fóruns dos jogadores de *League of Legends*® e eles vão muito além de meras adaptações de palavras de origem inglesa ou alternâncias que são realizadas, elas têm uma funcionalidade inerente ao ambiente. Outro fator a ser ponderado refere-se à Sociolinguística, pois podemos observar que existe uma relação entre as línguas que estão em contato dentro dessa comunidade.

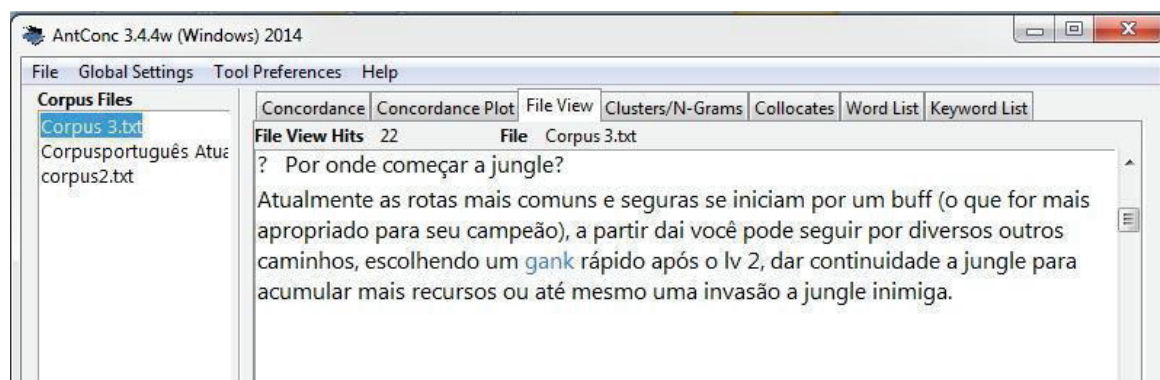
Temos, de um lado, a língua de prestígio (inglês), que tende a suplantar língua materna desprestigiada (português). Vemos que essa abordagem foi apontada e discutida por Muysken (2001), mesmo que em termos muito brandos e breves, em que os níveis de *CM* de acordo com as comunidades bilíngues estão interligados com o prestígio das línguas que serão alternadas, podendo interferir nas escolhas linguísticas em determinados momentos do discurso.

Não podemos nos esquecer, ainda, que esses discursos se referem a um ambiente atípico, pois não existem de fato duas comunidades ou mais em contato, que possuem diferentes línguas, mas sim apenas uma comunidade, que fala a mesma língua, mas opta pelo uso de vocabulário em uma segunda língua.

Ao compilar o *corpus* que contém interações entre jogadores de *League of Legends*® pudemos ter a noção mais precisa de qual é a frequência de uso de palavras mescladas de língua inglesa e língua portuguesa, assim como notamos o uso de palavras de língua inglesa que não sofreram nenhuma alteração em seu radical e em seu uso dentro das falas dos jogadores.

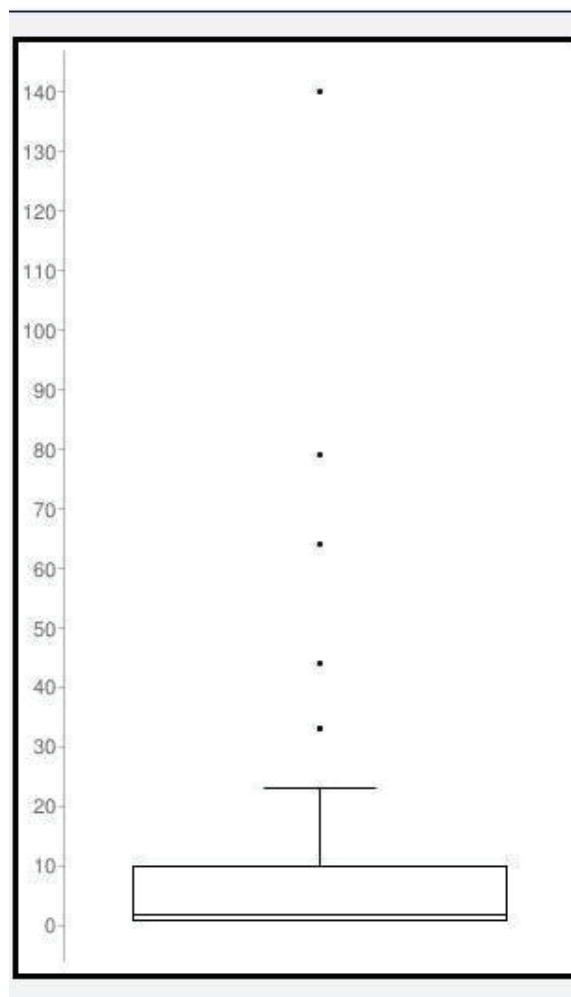
Podemos observar que, apesar de terem a frequência baixa em relação ao *corpus* inteiro destas misturas linguísticas, notamos que elas possuem importância nos comentários que são utilizados pelos jogadores, principalmente ao descrever ações ou nomear objetos, criaturas, habilidades etc., importantes para o jogador no contexto comunicativo, como podemos observar na figura abaixo:

Figura 21: Contexto de uso



Fonte: Captura de tela do programa *AntConc*® (Autor)

No excerto retirado do *corpus*, a palavra *gank* assume importância no uso do jogador, assim como podemos observar o uso de outros *CM*, como *buff* e *jungle*, permeando o seu discurso. Esse uso é recorrente e se repete ao longo dos textos extraídos, com maior ou menor intensidade. Ao relacionar essa intensidade de uso ao que o *corpus* apresenta em forma de frequência relativa, podemos ter uma melhor noção da razão probabilística do uso pelos jogadores. Assim, visamos apresentar, da maneira mais didática possível, um diagrama de caixa ou *boxplot*, que apresenta a média de uso dos *CM* e dos empréstimos, primeiramente referente aos *CM* e posteriormente aos empréstimos.

Figura 22: *Boxplot code-mixing*

Fonte: Elaborado pelo autor

No diagrama, vemos que existem os números de 0 a 140 e linhas que representam os quartis, estes quartis representam 25% cada, ou seja, de 0 até a primeira linha, possuímos um quartil (25%) das frequências das palavras coletadas, ou seja, um número pequeno de palavras possui pouca frequência no *corpus*, ou seja, poucas palavras têm frequência de 1 até 4; no entanto, palavras com frequências de 4 a 10 têm mais oportunidades de serem observadas no *corpus*.

Ao mesmo tempo, temos pontos que estão para fora do quartil, os quais são chamados de *outliers* ou “extrapolados”, pois são números que não cabem na média de palavras. No caso do diagrama referente aos *CM*, temos cinco palavras que são consideradas de limite extrapolado, as quais possuem uma alta incidência em relação às palavras restantes, como veremos no quadro a seguir:

Quadro 10: *Code-mixing* extrapolados

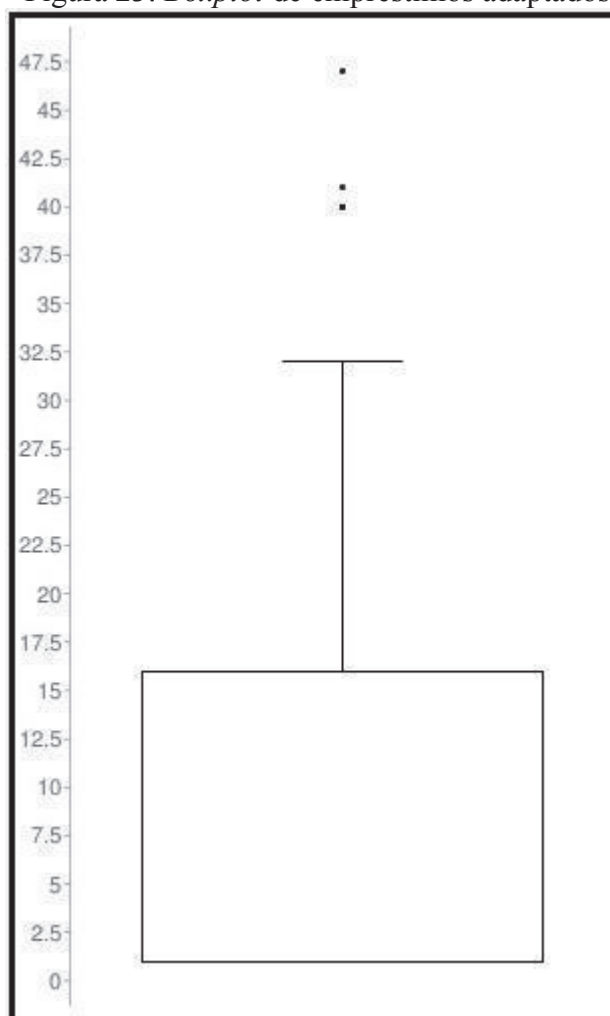
<i>Code-mixing</i>	Frequência	Total do <i>corpus</i>	Frequência (%)
<i>lane</i>	140	30442	0,45989094
<i>jungle</i>	79	30442	0,259509888
<i>game</i>	64	30442	0,210235858
<i>gank</i>	44	30442	0,144537153
<i>farm</i>	33	30442	0,108402864
<i>early</i>	33	30442	0,108402864

Fonte: Elaborado pelo autor

Podemos observar que os itens mais frequentes apresentados na compilação do *corpus* ocupam a função de substantivo, o que demonstra que a inserção de substantivos é mais frequente, corroborando, assim, a teoria apresentada por Muysken (2001) de que *CM* por inserção tendem a ter um número elevado de substantivos e palavras lexicais.

Temos, no topo do quadro, a palavra *lane*, seguida de *jungle*, *game*, *gank*, *farm* e, por último, um adjetivo, *early*, que tende a ser um item com certa regularidade nos fenômenos de inserção.

Em relação aos empréstimos que sofrem alguma adaptação morfológica ao longo do discurso dos jogadores, há uma conformidade maior de ocorrências, sendo a média mais homogênea, como poderemos observar na figura a seguir:

Figura 23: *Boxplot* de empréstimos adaptados

Fonte: Elaborado pelo autor

Esse diagrama refere-se à frequência dos empréstimos, no qual é possível perceber um quartil muito maior, que se estende do número 1 até o número 16 e estão relacionados às frequências de uso de empréstimos. Vemos que existe uma regularidade maior em relação aos *CM* por inserção no que se refere aos itens que possuem uma menor frequência, assim como os *CM*. Existem palavras que extrapolam a média, mas com um número menor, como veremos a seguir no quadro 12, referente aos empréstimos.

O quadro completo em relação aos números de frequência de cada item apresentado nos diagramas está alocado na seção Apêndice deste trabalho, sendo o Apêndice A referente aos *CM* e o B referente aos empréstimos que sofrem adaptações.

Observamos, ao longo deste trabalho, que a frequência de uso em relação aos *CM* que são substantivos é elevada. Observamos que, estatisticamente, os substantivos extrapolam o uso, sempre sendo muito mais frequentes em relação aos empréstimos adaptados, como vistos nas figuras 24 e 25.

Quadro 11: Empréstimos extrapolados

Empréstimos	Frequência	Total do <i>corpus</i>	Frequência (%)
<i>mid</i>	47	30442	0,154391958
<i>bot</i>	41	30442	0,134682347
<i>top</i>	40	30442	0,131397411
<i>jungler</i>	40	30442	0,131397411

Fonte: Autor

Os substantivos são amplamente utilizados nas falas dos jogadores ao fazerem seus comentários no fórum, assim como existe uma tendência a se utilizar menos verbos, porém mantendo-se uma média de utilização.

Outro aspecto se refere à correlação entre os itens que possuem maiores frequências, pois a palavra *lane* possui a maior frequência de uso em relação aos *CM*, e, no que se refere aos empréstimos, as palavras *mid*, *bot* e *top* têm frequências aproximadas, o que se deve a essas 4 palavras representarem os “caminhos” que os jogadores utilizam no jogo *League of Legends*®.

5 CONCLUSÃO E ENCAMINHAMENTOS

Apresentamos, ao longo deste trabalho, como os dados foram coletados, assim como as teorias que embasam esta pesquisa. Observamos que a LC tem servido como base para a exploração de *corpora* de tamanhos menores que podem apontar ocorrências mais específicas de determinadas palavras. Neste trabalho, a compilação de um *corpus* pequeno foi importante para observarmos a frequência de palavras de língua inglesa dentro de um contexto comunicativo totalmente em português, assim como o uso de adaptações morfológicas que emergiram no contexto comunicativo dessa comunidade virtual.

Os objetivos gerais foram discutidos e alcançados, pois a coleta do *corpus* realizada nesta pesquisa foi imprescindível, uma vez que, por meio dos dados coletados no fórum *online*, pode-se depreender que a linguagem dos jogadores tende a ser mesclada, inserindo-se palavras de língua inglesa em um contexto comunicativo de língua portuguesa, como se observa nos dados apresentados na seção “Análise”.

Em relação à pergunta que encabeçou a discussão para esta pesquisa - quais são os fenômenos relacionados às misturas de códigos linguísticos observados nas interações entre os jogadores neste fórum *online*? - vemos que os fenômenos apresentados se relacionaram aos CM por Inserção (MUYSKEN, 2001), sendo mais recorrentes os substantivos, com maior frequência nas interações dos jogadores. Muysken (2001) observou que essas categorias de mistura são recorrentes quando a comunicação é menos proficiente, sendo, geralmente, o primeiro tipo de mistura entre duas comunidades com códigos linguísticos diferentes. O que chama a atenção para essa Inserção é que os substantivos tendem a ser acompanhados por artigos definidos, que são mais frequentes, e artigos indefinidos, menos recorrentes. Outro aspecto sobre os CM está relacionado à posição das inserções que ocupam na frase, geralmente sendo utilizados como objeto da oração.

Outro fenômeno que foi levado em consideração refere-se aos empréstimos que são adaptados. O uso recorrente de sufixação “-ar” é aparente na comunicação entre os jogadores, que, na falta de uma palavra que não pode ser traduzida, tendem a utilizar o radical da língua inglesa e adaptá-la utilizando o referido sufixo nos verbos.

Nesse tipo de empréstimo, os verbos que foram adaptados também podem fazer parte de locuções verbais, como em “dar gank”, “tentar farmar”, “fazer poke”, o que pressupõe a existência de palavras da língua portuguesa que permitem essa forma de locução, sendo as mais frequentes com os verbos “tentar”, “dar” e “fazer”, relacionados a uma ação. Quando

não há uma locução, os empréstimos adaptados tendem a ser objetos da oração e, mesmo modificados, acabam ocupando outra posição na sentença.

Dos objetivos específicos, discutimos a compilação do *corpus* de interações entre jogadores de *League of Legends*® e sua análise, observando as misturas de códigos linguísticos, a fim de verificar quais as categorias de palavras que se misturam e como o fazem por meio da observação dos agrupamentos lexicais.

Já na seção de Metodologia, descrevemos o *corpus* coletado, que pode ser caracterizado como um *corpus* de tamanho pequeno, contendo 30.442 palavras e de uso específico para estudo, como abordado por Ghadessy, Henry e Roseberry (2001). Vemos que o tamanho e a especificidade do *corpus* podem evidenciar que, no nicho dos jogadores *online*, existe o uso de língua inglesa de maneira recorrente, assim como a transformação de palavras de radicais estrangeiros em palavras que são compreendidas pelos jogadores, algo que os *corpora* grandes podem não evidenciar.

Por meio da coleta do *corpus*, pudemos realizar a análise apresentada ao longo deste trabalho. Observamos que os *code-mixing* observados no *corpus* são relativamente mais frequentes que os empréstimos adaptados para a língua portuguesa e representam uma porcentagem maior no montante de palavras coletado no *corpus*. Em relação à segunda pergunta desta pesquisa - quais são as recorrências de uso desses fenômenos e o que eles nos dizem sobre a linguagem no contexto da interação entre os jogadores e usuários do fórum *online*? -, a análise demonstrou que os substantivos são mais representativos entre os *CM*, enquanto os verbos são os mais adaptados nos empréstimos e tendem a sofrer sufixação, sendo declinados de acordo com a pessoa como nos exemplos: “*Você pode gankar seu aliado*” e “*Warda as rotas da jungle*”, nos quais os verbos adaptados são declinados na segunda pessoa do singular.

Outro aspecto observado em relação à frequência e representado nos diagramas de caixa refere-se às disparidades entre as palavras utilizadas. Os substantivos extrapolam a média de uso por serem mais facilmente “encaixados” nas sentenças dos jogadores, remetendo aos usos que Muysken (2001) aborda de que, em ambientes de integração e interação de duas línguas, se tem inicialmente a inserção de palavras de conteúdo gramatical.

Em relação aos usos de palavras adaptadas da língua inglesa para a portuguesa, o uso de verbos e locuções verbais se apresenta como um fenômeno recorrente no uso dos empréstimos, assim, nota-se que a língua portuguesa permite tais alocações para empréstimos adaptados, mas o que se mais evidencia é a naturalidade com que o público pesquisado faz adaptações que se tornam comuns aos ambientes *online* e que são recorrentes dentro das

comunidades de jogadores e usuários dos jogos eletrônicos *online*. Esta pesquisa ainda teve como uma proposta futura observar outras comunidades de jogadores de outros *games* eletrônicos e constituir um *corpus* que seja específico dos fenômenos de mixagem e adaptações linguísticas, procurando por recorrências ou presença de outros *CM*. Acreditamos que a observação mais ampla e posterior comparação com textos produzidos por outras comunidades de jogadores brasileiros e internacionais podem nos apontar uma possível existência de dialetos próprios inerentes a esse universo.

Para tanto, sugerimos: (i) a expansão do *corpus* compilado nesta pesquisa para um *corpus* mais abrangente, abarcando comunidades de outros jogos eletrônicos *online*, brasileiras e estrangeiras; (ii) a busca de recorrências dentro dessas comunidades de jogadores, observando os agrupamentos lexicais, *CM* e empréstimos adaptados; (iii) o emprego dessas recorrências e vocabulário como base para a criação de uma plataforma digital que poderá ser utilizada por tradutores ou para pesquisa relacionada às variações linguísticas. A tecnologia informacional vem se solidificando cada vez mais nos ambientes comunicacionais, mudando e alterando a forma como diferentes linguagens podem estar em contato. Podemos ter, nesses ambientes de comunicação *online*, formas embrionárias de como duas línguas se influenciam e como o prestígio exerce influência sobre a língua desprestigiada, porém, deixamos essa observação para ser discutida posteriormente.

REFERÊNCIAS

- ANTHONY, L. **Lawrence Anthony Website (*AntConc*)**. Disponível em: <http://www.antlab.sci.waseda.ac.jp/index.html> Acesso em 01 mai. 2019.
- APPEL, R; MUYSKEN, P. **Language Contact and Bilingualism**. Amsterdam: Amsterdam University Press, p. 227, 2005.
- APPEL, R; MUYSKEN, P. **Code-switching and code-mixing**. In APPEL, R; MUYSKEN, P. **Language Contact and Bilingualism**. Amsterdam: Amsterdam University Press, p. 117-128, 2005.
- AUER, P. **Code-Switching In Conversnsation**. London: Routledge, 1998.
- BAKER, C. **Foundations of Bilingual Education and Bilingualism**. 3. ed. Clevedon: Biddles Ltd, 2001.
- BASILIO, M. **Formação e classes de palavras no português do Brasil**. 3. ed. São Paulo: Contexto, 2004.
- BECHARA, E. **Moderna Gramática Portuguesa**. 37. ed. Rio de Janeiro: Editora Nova Fronteira, 2009.
- BERBER SARDINHA, T. **Linguística De *Corpus*: Histórico E Problemática**. DELTA., São Paulo, v. 16, n. 2, 323-267, 2000.
- BERBER SARDINHA, T. **Corpora Eletrônicos Na Pesquisa Em Tradução**. Cadernos de Tradução, Florianópolis, v. 1, n. 9, 15-59, 2002.
- BERBER SARDINHA, T. **Linguística de *Corpus***. Barueri, São Paulo: Manole, 2004.
- BIBER, D.; CONRAD, S.; REPPEN, R. **Investigating Language Structure and Use**. Cambridge University Press, 1998.
- BLOM, J. P.; & GUMPERZ J. **Social meaning in linguistic structure: Code-switching in Norway**. In: Gumperz, J. P; Hymes, D. (Org.) *Directions in sociolinguistics*. New York: Holt, Rinehart and Winston, 1972.
- CACCAMO, C. A. **From ‘Switching Code’ To ‘Code-Switching’**. In: AUER, P. **Code-Switching In Conversnsation**. London: Routledge, 1998. p. 29 - 48.
- CÁRDENAS-CLAROS, M. S.; ISHARYANTI, N. **Code switching and code mixing in Internet chatting: between ‘yes’, ‘ya’, and ‘si’ a case study**. The Jaltcalljournal, Tokyo, v. 5, n. 3, p. 67-78, Jan. 2009.
- CLYNE, M. **Perspectives on Language Contact**. Hawthorn Press, Melbourne, v.1, p. 138, 1972.

FIORIN, J. L. **A internet vai acabar com a língua portuguesa?** Texto Livre: Linguagem e Tecnologia, Belo Horizonte - MG, v. 1, n. 1, p. 2-9, 2008.

GARDNER-CHLOROS, P. **Code-switching**. 1. ed. Nova York: Cambridge University Press, 2009.

GHADESSY, M.; HENRY, A; ROSEBERRY, R. L. **Small Corpus Studies and ELT: Studies and practice**. John Benjamins Publishing Company, Amsterdam, v.5, 2001.

GROSJEAN, J. **Studying Bilinguals**. 1ª ed. Oxford: Oxford Press, 2008.

HOFFER, L. B. **Language Borrowing and Language Diffusion: an Overview**. Intercultural Communication Studies, Trinity University. XI, p. 1-37, 2002.

HOFFER, L. B. **Language Borrowing and the Indices of Adaptability and Receptivity**. Intercultural Communication Studies, Trinity University. XVI: 2, p. 53-72, 2006.

IGN. **IGN Brasil**. Disponível em: <https://br.ign.com/>>. Acesso em: 16 mai. 2018.

LEAGUE OF LEGENDS. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2020. Disponível em: https://pt.wikipedia.org/w/index.php?title=League_of_Legends&oldid=58336722>. Acesso em: 24 mai. 2020.

MCENERY, T.; WILSON, A. **Corpus Linguistics: an introduction**. 2. ed. Edinburgh: Edinburgh University Press, 2001.

MUYSKEN, P.; SCIULLO, A. M.; SINGH, R. **Government and code-mixing**. Journal of Linguistics, Montreal, v.22, p.1-22, 1986.

MUYSKEN, P. **Bilingual Speech: A typology of code-mixing**. Cambridge: Cambridge University Press, 2001.

MYERS-SCOTTON, C. **Comparing codeswitching and borrowing**. Journal of Multilingual and Multicultural Development, Columbia, v. 13, n. 1, p. 19-39, 1992.

RIBEIRO, G. C. B. **Tradução Técnica, Terminologia E Linguística De Corpus: A Ferramenta Wordsmith Tools**. Cadernos de Tradução, Florianópolis, v. 2, n. 14, p. 159-174, 2004.

SINCLAIR, J. **Corpus, concordance, collocation: Describing English language**. Oxford: Oxford University Press, 1991.

LONGMAN DICTIONARY. **Longman Dictionary of Contemporary English**. 2 ed. Pearson Education, 2009. p. 769.

PFAFF, C. W. **Constraints on Language Mixing: Intrasentential Code-Switching and Borrowing in Spanish/English**. Language. Estados Unidos, v. 55, n. 2, Jun. 1979. p. 291-318

POPLACK, S.; MEECHAN, M. **How languages fit together in Codemixing**. International Journal of Bilingualism. Ottawa, v. 2, n. 2, p. 127-138, julho 1998.

POPLACK, S. **Sometimes I'll start a sentence in English Y TERMINO EN ESPAÑOL**. Linguistics, Nova York, v. 1, n. 1, p.581-618, 1980.

POPLACK, S. **Bilingualism and the vernacular**. In: HARTFORD, B; FOSTER, C. H.; VALDMAN, A. (Org). **Issues in international bilingual education: The role of the Vernacular**. Nova York: Plenum Press, p. 1-23, 1982.

POPLACK, S. **What does the Nonce Borrowing Hypothesis Hypothesize?** Cambridge University Press, Cambridge, v 15, n. 3, p. 644-648, 2011.

POPLACK, S. SANKOFF, D. **Code-Switching**. In: AMMON, U; DITTMAR, N; MATTHEIER, K. J.; TRUDGILL, P. **Sociolinguistics: An International Handbook of Science of Language and Society**: Berlin, New York, v 1, n 2, p. 589 – 596, jan. 2004.

RIOT. **Riot Games**. Disponível em: <https://www.riotgames.com/en>. Acesso em: 12 abr. 2020.

UGOT, M. **Language Choice, Code-Switching and Codemixing in Biase**. Global Journal Of Humanities, Ningeria, v. 8, n. 2, p. 27-35, ago. 2010.

WEISSER, M. **Practical Corpus Linguistics**. The United Kingdom: John Wiley & Sons ltd, 2016.

WEINRICH, U. **Languages in contact: Findings and Problems**. 6. ed. The Hague: Mouton Publishers. 1968.

WEI, L. **The ‘why’ and ‘how’ questions in the analysis of conversational code switching**. In: AUER, P. (Org). **Code switching in conversation: Language, interaction and identity**. Londres: Routledge, 1998. p. 156-176.

APÊNDICE

APÊNDICE A - Frequência dos *code-mixing* observados no *corpus*

Code-mixing	Frequência	Total do <i>corpus</i>		%	Frequência (%)
lane	140	30442	0,00459 9	100	0,45989094
jungle	79	30442	0,00259 5	100	0,25950988 8
game	64	30442	0,00210 2	100	0,21023585 8
gank	44	30442	0,00144 5	100	0,14453715 3
farm	33	30442	0,00108 4	100	0,10840286 4
early	33	30442	0,00108 4	100	0,10840286 4
kill	23	30442	0,00075 6	100	0,07555351 2
skills	20	30442	0,00065 7	100	0,06569870 6
ward	17	30442	0,00055 8	100	0,0558439
tank	15	30442	0,00049 3	100	0,04927402 9
hit	15	30442	0,00049 3	100	0,04927402 9
brush	15	30442	0,00049 3	100	0,04927402 9
carry	15	30442	0,00049 3	100	0,04927402 9
gold	13	30442	0,00042 7	100	0,04270415 9
burst	12	30442	0,00039 4	100	0,03941922 3
post	12	30442	0,00039 4	100	0,03941922 3
push	12	30442	0,00039 4	100	0,03941922 3
level	12	30442	0,00039 4	100	0,03941922 3
champions	12	30442	0,00039 4	100	0,03941922 3
skill	11	30442	0,00036	100	0,03613428

			1		8
poke	11	30442	0,00036 1	100	0,03613428 8
split	11	30442	0,00036 1	100	0,03613428 8
team	11	30442	0,00036 1	100	0,03613428 8
champion	10	30442	0,00032 8	100	0,03284935 3
main	10	30442	0,00032 8	100	0,03284935 3
speed	10	30442	0,00032 8	100	0,03284935 3
graves	10	30442	0,00032 8	100	0,03284935 3
sin	9	30442	0,00029 6	100	0,02956441 8
baron	9	30442	0,00029 6	100	0,02956441 8
masteries	9	30442	0,00029 6	100	0,02956441 8
attack	8	30442	0,00026 3	100	0,02627948 2
cooldown	8	30442	0,00026 3	100	0,02627948 2
dive	8	30442	0,00026 3	100	0,02627948 2
escape	8	30442	0,00026 3	100	0,02627948 2
pots	8	30442	0,00026 3	100	0,02627948 2
waves	8	30442	0,00026 3	100	0,02627948 2
miss	7	30442	0,00023	100	0,02299454 7
rift	7	30442	0,00023	100	0,02299454 7
smite	7	30442	0,00023	100	0,02299454 7
spell	6	30442	0,00019 7	100	0,01970961 2
teamfight	6	30442	0,00019 7	100	0,01970961 2
core	6	30442	0,00019 7	100	0,01970961 2
snowball	5	30442	0,00016	100	0,01642467

			4		6
master	4	30442	0,00013 1	100	0,01313974 1
wave	4	30442	0,00013 1	100	0,01313974 1
hard	4	30442	0,00013 1	100	0,01313974 1
pokes	3	30442	9,85E-05	100	0,00985480 6
middle	3	30442	9,85E-05	100	0,00985480 6
pot	2	30442	6,57E-05	100	0,00656987 1
stun	2	30442	6,57E-05	100	0,00656987 1
skin	2	30442	6,57E-05	100	0,00656987 1
grave	1	30442	3,28E-05	100	0,00328493 5
squish	1	30442	3,28E-05	100	0,00328493 5
rank	1	30442	3,28E-05	100	0,00328493 5
stomp	1	30442	3,28E-05	100	0,00328493 5
backline	1	30442	3,28E-05	100	0,00328493 5
guardian	1	30442	3,28E-05	100	0,00328493 5
guide	1	30442	3,28E-05	100	0,00328493 5
hypercarry	1	30442	3,28E-05	100	0,00328493 5
junk	1	30442	3,28E-05	100	0,00328493 5
knock	1	30442	3,28E-05	100	0,00328493 5
launcher	1	30442	3,28E-05	100	0,00328493 5
like	1	30442	3,28E-05	100	0,00328493 5
line	1	30442	3,28E-05	100	0,00328493 5
medler	1	30442	3,28E-05	100	0,00328493 5
move	1	30442	3,28E-05	100	0,00328493

					5
oracle	1	30442	3,28E-05	100	0,003284935
out	1	30442	3,28E-05	100	0,003284935
outdamage	1	30442	3,28E-05	100	0,003284935
outfarm	1	30442	3,28E-05	100	0,003284935
outrange	1	30442	3,28E-05	100	0,003284935
phantom	1	30442	3,28E-05	100	0,003284935
pit	1	30442	3,28E-05	100	0,003284935
ranger	1	30442	3,28E-05	100	0,003284935
reset	1	30442	3,28E-05	100	0,003284935
rework	1	30442	3,28E-05	100	0,003284935
save	1	30442	3,28E-05	100	0,003284935
scaling	1	30442	3,28E-05	100	0,003284935
scape	1	30442	3,28E-05	100	0,003284935
score	1	30442	3,28E-05	100	0,003284935
screenshot	1	30442	3,28E-05	100	0,003284935
single	1	30442	3,28E-05	100	0,003284935
slash	1	30442	3,28E-05	100	0,003284935
smart	1	30442	3,28E-05	100	0,003284935
smoke	1	30442	3,28E-05	100	0,003284935
snare	1	30442	3,28E-05	100	0,003284935
spike	1	30442	3,28E-05	100	0,003284935
spin	1	30442	3,28E-05	100	0,003284935
squisy	1	30442	3,28E-05	100	0,003284935

					5
support	1	30442	3,28E-05	100	0,003284935
target	1	30442	3,28E-05	100	0,003284935
teambattles	1	30442	3,28E-05	100	0,003284935
hits	1	30442	3,28E-05	100	0,003284935
spells	1	30442	3,28E-05	100	0,003284935
skins	1	30442	3,28E-05	100	0,003284935
stuns	1	30442	3,28E-05	100	0,003284935
sinful	1	30442	3,28E-05	100	0,003284935
teamfights	1	30442	3,28E-05	100	0,003284935
attacks	1	30442	3,28E-05	100	0,003284935
squishy	1	30442	3,28E-05	100	0,003284935
tanks	1	30442	3,28E-05	100	0,003284935

APÊNDICE B - Frequência dos empréstimos adaptados observados no *corpus*

Empréstimos	Frequência	Total do <i>corpus</i>		%	Frequência (%)
mid	47	30442	0,001544	100	0,154391958
bot	41	30442	0,001347	100	0,134682347
top	40	30442	0,001314	100	0,131397411
jungler	40	30442	0,001314	100	0,131397411
ult	32	30442	0,001051	100	0,105117929
buff	24	30442	0,000788	100	0,078838447
laner	23	30442	0,000756	100	0,075553512
ganks	22	30442	0,000723	100	0,072268576
ashe	20	30442	0,000657	100	0,065698706
farmar	16	30442	0,000526	100	0,052558965
lanes	16	30442	0,000526	100	0,052558965
buffs	15	30442	0,000493	100	0,049274029
gankar	12	30442	0,000394	100	0,039419223
laners	9	30442	0,000296	100	0,029564418
skillshot	8	30442	0,000263	100	0,026279482
ultar	7	30442	0,00023	100	0,022994547
wardar	7	30442	0,00023	100	0,022994547
logar	7	30442	0,00023	100	0,022994547
botlane	1	30442	3,28E-05	100	0,003284935
junglar	1	30442	3,28E-05	100	0,003284935
leveis	1	30442	3,28E-05	100	0,003284935
stunar	1	30442	3,28E-05	100	0,003284935
pentakill	1	30442	3,28E-05	100	0,003284935
offtank	1	30442	3,28E-05	100	0,003284935
tanka	1	30442	3,28E-05	100	0,003284935
snowballar	1	30442	3,28E-05	100	0,003284935
stunou	1	30442	3,28E-05	100	0,003284935
midlaner	1	30442	3,28E-05	100	0,003284935
stackado	1	30442	3,28E-05	100	0,003284935
toplane	1	30442	3,28E-05	100	0,003284935
masterizei	1	30442	3,28E-05	100	0,003284935
squishes	1	30442	3,28E-05	100	0,003284935
midlanner	1	30442	3,28E-05	100	0,003284935
ranqueada	1	30442	3,28E-05	100	0,003284935
stackar	1	30442	3,28E-05	100	0,003284935
stompar	1	30442	3,28E-05	100	0,003284935
gankou	1	30442	3,28E-05	100	0,003284935
wardaram	1	30442	3,28E-05	100	0,003284935
tankers	1	30442	3,28E-05	100	0,003284935

wardou	1	30442	3,28E-05	100	0,003284935
farmou	1	30442	3,28E-05	100	0,003284935