



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE LETRAS – IL
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO – LET
LÍNGUAS ESTRANGEIRAS APLICADAS – LEA-MSI

Vítor Amaral Córdoba

CRIAÇÃO DE UM CORPUS PARALELO COREANO-PORTUGUÊS-INGLÊS DE
ESPORTS

BRASÍLIA, DF

2023

Vítor Amaral Córdoba

**CRIAÇÃO DE UM CORPUS PARALELO COREANO-PORTUGUÊS-INGLÊS DE
*ESPORTS***

Trabalho de Conclusão de Curso (TCC) apresentado ao departamento de Línguas Estrangeiras e Tradução da Universidade de Brasília, como requisito parcial à obtenção do título de Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI).

Orientador: Prof. Dr. Thiago Blanch Pires

BRASÍLIA, DF

2023

Vítor Amaral Córdoba

**CRIAÇÃO DE UM CORPUS PARALELO COREANO-PORTUGUÊS-INGLÊS DE
*ESPORTS***

Trabalho de Conclusão de Curso (TCC) apresentado ao departamento de Línguas Estrangeiras e Tradução da Universidade de Brasília, como requisito parcial à obtenção do título de Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI).

Brasília, 18 de dezembro de 2023

BANCA EXAMINADORA

Prof. Dr. Thiago Blanch Pires
(Orientador – LET/UnB)

Profa. Dra. Clarissa Prado Marini
(Avaliadora – LET/UnB)

Prof. Dr. Marcus Vinicius de Lira Ferreira Tanaka
(Avaliador – LET/UnB)

AGRADECIMENTOS

Sempre pensei que a seção de agradecimentos é ao mesmo tempo um registro temporal de como estavam as coisas quando escrevi a mensagem e um registro injusto por sua finitude de nomes, deixando de fora muitos que são esquecidos no momento de escrita. Mesmo assim, seria pior ainda deixar passar em branco a oportunidade de agradecer pessoas tão importantes.

Agradeço ao meu pai pelos conselhos e por sempre procurar entender sobre o que estou falando ou fazendo;

à minha mãe e à minha irmã, pelas conversas durante a janta e por serem fontes de respostas para todas as minhas perguntas “burras”;

à minha vó Josina, por sempre se dedicar inteiramente aos seus netos;

ao vovô Zezinho e à vovó Cidalica, por servirem de exemplo de amor e carinho;

aos meus primos: Biel, Cacala, Clarice, Gui, Juliana e Luciana, que foram meus primeiros amigos e que seguem me apoiando, mesmo hoje em dia com rotinas tão diferentes;

à Débora, por estar sempre ao meu lado e por me ajudar incondicionalmente;

aos meus amigos: André, Artur, Davi, Dota, Filipão, Ramon, Taboada e Vitor, com quem passo quase todos os meus dias com;

e a todos àqueles que não tive a felicidade de me recordar nos 5 minutos que dediquei a este trecho, mas que com certeza fizeram/fazem parte da minha história e me impactaram para que eu pudesse me tornar quem sou.

*Aquele que trabalha duro pode superar um gênio.
Mas de nada adianta trabalho duro se você não
confia em você mesmo.”*

Rock Lee

RESUMO

Apesar da Coreia do Sul ser um país com vasta história e influência no âmbito dos esportes eletrônicos – *esports*, é evidente que a barreira linguística da tradução impede uma maior interação do conteúdo coreano com o resto do mundo. Este trabalho busca introduzir a importância da Coreia do Sul no âmbito dos *esports* e elaborar um corpus paralelo coreano-português-inglês com enfoque no domínio dos *esports*, mais especificamente no jogo *League of Legends*. O corpus foi montado a partir das documentações oficiais do jogos e da coleta de transcrições de falas de jogadores profissionais da liga coreana (LCK) durante jogos oficiais e transmissões de partidas online, a fim de incluir tanto a linguagem oficial presente nos arquivos do jogo quanto a linguagem realisticamente utilizada pela comunidade de jogadores. Para a coleta de dados exclusivos da língua coreana ou com tradução somente para o inglês, foi realizada a tradução manual para o português. O corpus conta com 36128 *types* e 297684 *tokens* e espera-se que a sua criação contribua com a cooperação bilateral sul-coreana e brasileira, e que torne o conteúdo dos *esports* sul-coreanos mais acessível e compreensível para a audiência brasileira. O corpus tem como objetivo introduzir a linguagem específica de jogos do coreano à comunidade brasileira em seu contexto de uso real e servir como um corpus de referência em uma área de conteúdo escasso, qual seja, a da tradução de conteúdo de jogos eletrônicos do coreano para o português.

Palavras-chave: *esports*; linguística de corpus; coreano; terminologia

ABSTRACT

Although South Korea is a country with a vast history and influence in the field of electronic sports – *esports*, the language translation barrier prevents greater interaction between Korean content and the rest of the world. This work seeks to introduce the importance of South Korea in the context of esports and develop a parallel Korean-Portuguese-English corpus focusing on the domain of esports, more specifically on the game League of Legends. The corpus was assembled from official game documentation and the collection of transcriptions of speeches from professional players from the Korean league (LCK) during official games and online broadcasts, to include both the official language present in the game files and the realistic language used by the gaming community. To collect data present only in Korean or with translation only into English, manual translation was made to Portuguese. The corpus has 36,128 *types* and 297,684 *tokens* and its creation is expected to contribute to South Korean and Brazilian bilateral cooperation, and to make South Korean esports content more accessible and understandable for the Brazilian audience. The corpus aims to introduce the specific language of Korean games to the Brazilian community in its context of real use and to serve as a reference corpus in an area of scarce content, namely, the translation of video game contents from Korean to Portuguese.

Keywords: esports; corpus linguistics; Korean; terminology

초록

한국은 e 스포츠 분야에서 유구한 역사와 영향력을 가진 국가이지만 언어의 장벽으로 인해 그 영향력이 전 세계적으로 확산되는데 어려움을 겪고 있다. 본 연구는 e 스포츠 분야에서 한국의 중요성을 소개하고, 한국어-포르투갈어-영어를 병행한 e 스포츠 코퍼스를 제시하고자 한다. 해당 코퍼스는 e 스포츠 중에서도 게임 '리그 오브 레전드'에 초점을 맞췄다. 코퍼스 구축에 사용된 자료는 공식 게임 문서를 비롯해 실제 경기 및 한국 리그(LCK) 선수들의 온라인 방송에서 발화된 내용을 수집하였고, 게임 내에서 공식적으로 사용되는 언어 뿐만 아니라 게임 커뮤니티에서 사용되는 실제적인 언어까지 모두 포함한다. 데이터를 수집하는 과정에서 한국어만 제시되거나 영어로만 번역된 경우에는 포르투갈어로 직접 번역을 진행했다. 본 코퍼스는 총 36,128 개의 유형과 297,684 개의 토큰을 바탕으로 구축되었으며, 실제로 사용되는 한국 게임 용어들을 브라질 커뮤니티에 소개하고, 브라질 e 스포츠 콘텐츠 부족을 보완하는 과정에서 한국어 e 스포츠 콘텐츠를 포르투갈어로 번역할 때 참조 역할을 하는 것을 목표로 한다. 이를 통해 브라질 관객들이 한국의 e 스포츠 콘텐츠를 보다 쉽고 원활하게 이해하고, 더 나아가 한국과 브라질의 양자 협력으로도 이어지기를 바란다.

키워드: e 스포츠, 말뭉치 언어학, 한국어, 용어

LISTA DE FIGURAS

Figura 1 – Estruturação do <i>English-Norwegian Parallel Corpus (ENPC)</i>	24
Figura 2 – Estruturação do <i>Oslo Multilingual Corpus (OMC)</i>	25
Figura 3 – Estruturação do corpus paralelo coreano-português-inglês no domínio de <i>esports</i>	30
Figura 4 – exemplo de arquivo .json em inglês do campeão Aatrox	32
Figura 5 – captura de tela de vídeo da T1	33
Figura 6 – captura de tela de legenda de vídeo da T1	34
Figura 7 – transcrição realizada pelo <i>Tesseract</i>	34
Figura 8 – página inicial do software <i>VideoSubFinder</i>	35
Figura 9 – janela de configurações do <i>VideoSubFinder</i>	36
Figura 10 – imagem base para processamento no <i>VideoSubFinder</i>	36
Figura 11 – imagem com processamento de filtro de cores inicial	37
Figura 12 – imagem com processamento de filtro de cores com bordas.....	37
Figura 13 – imagem com processamento de cores final.....	38
Figura 14 – pastas de arquivos de imagem geradas pelo <i>VideoSubFinder</i>	38
Figura 15 – exemplo imagem final da pasta <i>TX/Images</i>	39
Figura 16 – recorte de arquivo único da pasta <i>ImagesJoined</i>	39
Figura 17 – arquivo após OCR realizado pelo Google Docs	40
Figura 18 – função no Python para remoção de <i>stopwords</i> no ranqueamento de palavras.....	43

LISTA DE GRÁFICOS

Gráfico 1 – Número de jogadores sul-coreanos participantes do Campeonato Mundial por liga 18

LISTA DE TABELAS

Tabela 1 – Servidores geridos pela Riot Games.....	17
Tabela 2 – coleção de vídeos subcorpus 2.....	40
Tabela 3 – coleção de entrevistas do subcorpus 3.....	41
Tabela 4 – número de tokens e types do corpus.....	43
Tabela 5 – número de tokens e types subcorpus 1.....	44
Tabela 6 – 25 types mais frequentes subcorpus 1.....	45
Tabela 7 – número de tokens e types subcorpus 2.....	47
Tabela 8 – 25 types mais frequentes subcorpus 2.....	48
Tabela 9 – número de tokens e types subcorpus 3.....	49
Tabela 10 – 25 types mais frequentes subcorpus 3.....	50
Tabela 11 – equivalência de termos do subcorpus 3.....	51
Tabela 12 – abreviações no subcorpus 2.....	53
Tabela 13 – termos similares entre as três línguas.....	56

SUMÁRIO

1.	INTRODUÇÃO.....	8
2.	CONTEXTO DE INVESTIGAÇÃO	9
2.1.	O TERMO “ <i>ESPORTS</i> ”	9
2.2.	ORIGEM DOS <i>ESPORTS</i>	10
2.3.	ORIGEM DOS <i>ESPORTS</i> NA COREIA DO SUL	11
2.4.	O SURGIMENTO DO <i>LEAGUE OF LEGENDS</i>	13
2.5.	O SUCESSO MUNDIAL DA COREIA DO SUL.....	14
2.6.	LIMITAÇÕES PARA A DIFUSÃO DO CONTEÚDO COREANO A OUTRAS COMUNIDADES	19
3.	FUNDAMENTAÇÃO TEÓRICA	21
3.1.	LINGUÍSTICA DE CORPUS	21
3.2.	CORPUS PARALELO	22
3.3.	ESTUDOS DE TRADUÇÃO	23
3.4.	ESTRUTURAÇÃO DO CORPUS	24
3.5.	TERMINOLOGIA – TGT E TCT	26
3.6.	TERMINOLOGIA NO CORPUS	27
4.	METODOLOGIA.....	29
4.1.	SUBCORPUS 1 – DOCUMENTAÇÕES DE JOGOS.....	31
4.2.	SUBCORPUS 2 – TRANSCRIÇÕES DA COMUNICAÇÃO DE JOGADORES PROFISSIONAIS EM PARTIDAS OFICIAIS E DAS FALAS DOS JOGADORES EM TRANSMISSÕES AO VIVO	32
4.3.	SUBCORPUS 3 – ENTREVISTAS REALIZADAS COM JOGADORES PROFISSIONAIS.....	41
4.4.	COMPILAÇÃO E ANÁLISE DO CORPUS	42
5.	RESULTADOS	43
5.1.	ANÁLISE INTRACORPUS	44
5.2.	ANÁLISE INTERCORPORA.....	52
6.	CONSIDERAÇÕES FINAIS	56

1. INTRODUÇÃO

Os esportes eletrônicos, amplamente conhecidos como *esports*, surgiram como uma forma de entretenimento e competição de rápido crescimento nas últimas décadas. Essa indústria diversificada e moderna, que abrange uma variedade de videogames competitivos, conquistou uma audiência global relevante. Com eventos lotados, transmissões ao vivo que atraem milhões de espectadores e atletas profissionais ganhando reconhecimento internacional, os *esports* consolidaram sua posição como uma parte vital da cultura contemporânea e são responsáveis por fomentar a indústria dos videogames, uma das maiores do mundo.

No epicentro desse fenômeno, encontra-se a Coreia do Sul, um país que se destaca como um dos maiores destaques mundiais dos *esports*. Com uma rica história de competições de videogames e um rápido desenvolvimento tecnológico, a Coreia do Sul tornou-se um bastião inegável dos *esports*, influenciando a indústria global de maneira inestimável. No entanto, apesar do sucesso dos *esports* sul-coreanos, uma barreira persistente permanece: a barreira linguística.

É nesse contexto que esta pesquisa se insere. Embora os sul-coreanos tenham uma influência notável nos *esports*, o acesso a conteúdo relevante e informações detalhadas sobre esse mundo muitas vezes se mostra inacessível para aqueles que não falam coreano. Os termos específicos e as abreviações utilizadas pelas comunidades de *esports* sul-coreanas frequentemente constituem uma linguagem de especialidade, obscurecendo ainda mais o entendimento para aqueles que não estão familiarizados com ela.

É com base neste contexto que o presente estudo aborda a escassez de tradução de termos do coreano para outras línguas nos *esports* como seu problema central. Assim, o objetivo geral desta investigação é elaborar um corpus paralelo coreano-português-inglês de *esports*, com enfoque em trazer também o português além do inglês devido a apenas uma pequena parcela da população brasileira ser proficiente na língua inglesa. O corpus incluirá os termos usados pelos falantes coreanos em jogos online, com destaque no jogo *League of Legends*, devido ao seu impacto e influência no meio dos *esports*. Isso permitirá uma compreensão mais profunda e acessível desse cenário de *esports* sul-coreanos para a audiência brasileira, além de servir como um corpus de referência em uma área de conteúdo escasso.

Inicialmente, será brevemente introduzida a história dos *esports* e motivos para a escolha de sua grafia ser o hífen ou capitalização no capítulo 2. Em seguida, ainda no mesmo capítulo, uma breve introdução ao desenvolvimento tecnológico coreano e à sua cultura de *PC bangs*, e

o paralelo com a realidade brasileira, trazendo informações sobre o seu público e possíveis benefícios com o acesso a informações sul-coreanas.

Para os leitores que buscam apenas informações sobre o desenvolvimento do corpus e o seu referencial teórico, as informações diretamente relacionadas à criação do corpus serão introduzidas no capítulo 3, da fundamentação teórica.

Após a contextualização do problema e fundamentação teórica para a criação do corpus, descreveremos a metodologia utilizada para compilar um corpus paralelo coreano-português-inglês, que incluirá os termos utilizados pelos falantes coreanos em jogos online no capítulo 4.

Por fim, apresentaremos os resultados desta pesquisa no capítulo 5 e discutiremos possíveis aplicações do corpus e delinearemos planos futuros para a continuidade deste trabalho, destacando sua contribuição para tornar o conteúdo dos *esports* sul-coreanos mais acessível e compreensível para a audiência brasileira no capítulo 6.

2. CONTEXTO DE INVESTIGAÇÃO

2.1. O termo “*esports*”

Os esportes eletrônicos, comumente apelidados de *esports*, são um formato de competição esportiva em que sua aplicação ocorre por meio de sistemas eletrônicos (Hamari; Sjöblom, 2017). O termo começou a ser utilizado em meados dos anos 90, com uma das primeiras recordações de uso em veículos midiáticos relevantes em 1999, em publicação jornalística da página EuroGamer. A publicação citava a criação da associação esportiva *Online Gamers Association* (OGA) e a sua importância para a representação de jogadores profissionais de jogos eletrônicos para o desenvolvimento competitivo do cenário. Em entrevista com uma das figuras relevantes do projeto, Mat Bettinson compara os esportes eletrônicos com esportes tradicionais e faz um dos primeiros usos documentados do termo: “Certamente não demorará muito para que os eSports sejam cobertos pela televisão da mesma forma que os esportes tradicionais (Bettinson, 1999, tradução nossa)”

Desde a sua criação, o termo gerou discussões referentes à sua grafia e significado, como por exemplo a capitalização do “s” em *sports*, ou a inclusão de um hífen (e-sports), sob a premissa de seguir a padronização de e-mail, referente a correio eletrônico. Devido à divergência dos *esports* a esportes tradicionais e a sua ausência de características físicas, houve também resistência em sua aceitação como esporte, e não como passatempo ou apenas o ato de tornar jogos eletrônicos competitivos. Com o rápido desenvolvimento do cenário competitivo,

crescimento da indústria dos videogames e profissionalização do meio, seria uma visão negacionista não reconhecer os *esports* como esporte.

Em relação à grafia utilizada, a agência de notícias americana *Associated Press* publicou por meio de seu guia estilístico de que a grafia correta a ser utilizada por se referir a videogames multijogador competitivos, com a utilização de variantes apenas para se referir a nomes próprios (APStylebook, 2017). Referente ao uso do termo em português, o artigo jornalístico de Rafael Romer, “*eSports, esports* ou *e-sports*: Qual a grafia correta dos esportes eletrônicos?” (2019) traz um excelente resumo de seu histórico no país, e compara o termo com a palavra “e-mail”, que apesar de ter sua grafia com o hífen no vocabulário ortográfico da Academia Brasileira de Letras, a escrita sem o hífen, “email”, é comumente utilizada entre os falantes da língua, e no caso de “esports”, da mesma forma, a tendência é de que o hífen não seja utilizado (Romer, 2019).

2.2. Origem dos *esports*

Apesar de uma popularização maior do termo nos anos 90, a origem dos *esports* é incerta, com autores do tema afirmando que seu surgimento foi em 1947, com a introdução dos *flíperes* ao jogo do pinball, tornando o popular fliperama da época em um foco de competição e demonstração de habilidade em um bar no Reino Unido (Collis, 2020, p. 7). Outros afirmam que a sua origem se deu em 1972, com a criação de um torneio do jogo de computador *Spacewar* em Stanford denominado “*Intergalactic Spacewar Olympics*”, com a premiação para o vencedor de um ano de subscrição à revista *Rolling Stone* (Hiltscher, 2015).

Com a maior popularização dos videogames e crescente acesso à internet nos anos 90, os primeiros torneios online de maior notoriedade começaram a surgir, com destaque especial ao *Nintendo World Championships* em 1990, sucessor do *Nintendo Challenge Championship*, evento de grande escala sediado no Canadá em 1989. O evento canadense contou com mais de 80.000 participantes em todo o país, e devido ao seu sucesso, motivou a criação de um campeonato mundial – apesar de restrito ao Canadá e aos Estados Unidos – com os mais bem colocados de cada estado (Nintendo Power Flash, 1989). A divisão de categorias foi realizada por idade, e teve a sua final sediada em Los Angeles após meses de competição em mais de 29 estados norte-americanos. Empresas como a Nintendo, Blockbuster e Atari foram grandes responsáveis pela disseminação de campeonatos de videogame antes dos anos 2000, com maior enfoque no público norte-americano e europeu. Durante o mesmo período, a realidade era diferente para o público sul-coreano em relação aos videogames.

2.3. Origem dos *esports* na Coreia do Sul

A Coreia do Sul investiu fartamente no desenvolvimento de uma sociedade da informação por meio de políticas governamentais e investimento público nas décadas de 80 e 90. Kim (2014) traz em sua análise da obra de Oh e Larson (2011) um breve resumo sobre o que é o conceito de uma sociedade da informação:

A ideia de uma “sociedade da informação” é inspirada no trabalho de Manuel Castells. Segundo os autores, este conceito envolve cinco “abordagens”: criações tecnológicas, crescimento econômico, mudanças ocupacionais, melhorias drásticas nas velocidades e distâncias de comunicação e a crescente abundância e utilização da informação na sociedade (Kim, 2014, p. 5, tradução nossa).

Kim (2014) aponta que a justificativa de utilização do conceito de Castells existe para melhor compreender a “revolução da informação” que ocorreu no país na década de 1980 – um período crítico para a indústria da tecnologia da informação e comunicação (TIC) que visava o desenvolvimento de tecnologias produzidas em massa, com semicondutores e computadores cada vez mais avançados.

Um dos pilares dessa “revolução” foi a criação da *Korean Information Infrastructure* (KII), com o objetivo de disponibilizar internet de alta velocidade para todo o país. Rea (2016) aponta que provedores de internet coreanos participantes da KII alcançavam mais de 1.6 milhão de assinaturas em fevereiro de 1997. Apesar disso, o custo de manutenção das subscrições era tão alto que a expansão do projeto se mostrava desafiadora, o que levou à criação de centros de acesso comunitários. Inicialmente, a criação dos centros comunitários possibilitou o acesso à internet para pessoas que não teriam condições financeiras para tal feito de maneira privada em suas casas. Porém, com o final do ano de 1997 também trazendo uma das maiores crises financeiras da história da Ásia, os planos da KII foram afetados e a manutenção dos centros de acesso comunitários tornou-se inviável sob comando público.

Em momento oportuno em meio da severa crise, a iniciativa privada e a intervenção do FMI no país possibilitaram o desenvolvimento de projetos privados que buscavam transformar os centros de acesso comunitários em uma variante aos *internet cafés*, em que o usuário teria acesso a um computador com rápido acesso à internet por um pequeno valor a ser pago por hora utilizada. Esses estabelecimentos levavam o nome de *PC Bangs* – “bang”, da romanização da palavra coreana 방, pode ser traduzida como “quarto”, ou “sala”. Tal composição morfológica pode ser vista em outras iniciativas populares, como o *karaokê*, chamado na Coreia do Sul de *norebang* (*nore* (노래) significando “música” e *bang* (방) significando “sala” ou “quarto”).

A iniciativa foi um sucesso e seu crescimento pode ser comprovado em números, conforme atestado por Rea (2016):

“O número de PC bangs cresceu de cerca de 100 em 1997 para mais de 13.000 em 1999. O que impulsionou esta vertiginosa taxa de crescimento não foi tanto a disponibilidade de computadores para navegar na web ou verificar e-mails, mas sim o acesso barato e confiável a jogos online e às conexões de alta velocidade necessárias para jogá-los.” (Rea, 2016, p. 22, tradução nossa)

Desde então, a Coreia do Sul demonstrou imenso crescimento na área e é considerada por muitos como o país de maior influência global nos *esports*. Apelidada de “capital dos *esports*” (Yu, 2015), a nação sul-coreana foi uma das pioneiras no consumo do jogo StarCraft, um jogo de estratégia em tempo real (RTS) publicado pela desenvolvedora norte-americana *Blizzard*, em 1998. A complexidade do jogo e possibilidade de jogar com outros jogadores online tornou-se um sucesso na Coreia devido à acessibilidade ao jogo facilitada pelas PC Bangs, uma vez que mesmo que os indivíduos que não tinham recursos financeiros para comprar e jogar o jogo em casa, poderiam experimentar e praticar o jogo por preços razoáveis – valores abaixo de um dólar americano por hora – fora de casa. Começa então a formação de uma cultura PC bang (Jin, 2020, p. 3733), em que crianças, adolescentes e jovens adultos na Coreia do Sul tinham como norma o hobby de jogar videogames nas PC Bangs. Tal fenômeno é demonstrado por vários autores: “Kyung-joon, que estava no ensino fundamental em 1998, me disse que se um garoto de sua turma não soubesse jogar StarCraft, ele seria rotulado de *wang-tta* (perdedor) e rejeitado por seus colegas.” (Rea, 2016, p. 22, tradução nossa). “Na verdade, o StarCraft em conjunto com as PC Bangs mudou os padrões de vida dos jovens coreanos.” (Jin, 2020, tradução nossa).

Em dezembro de 1997, com a criação da *Korea Pro Gamers League* (KPGL) pela Hitel, o Starcraft ganhava espaço como um dos maiores jogos nos *esports* da década de 90: “Não é incorreto afirmar que o fenômeno do pleno andamento dos *esports* começou com StarCraft” (Jin, 2020, tradução nossa). O rápido desenvolvimento profissionalizado do meio foi incentivado por torneios realizados entre comunidades de jogadores e donos de PC Bangs que utilizavam do jogo e dos campeonatos para divulgar o seu comércio.

Conforme indica Rea (2016), as pessoas não estavam mais indo às PC bangs para jogar, mas também para assistirem outras pessoas jogando, e com isso pequenos campeonatos e competições foram se formando: “os melhores jogadores construía suas reputações na comunidade, desenvolviam rivalidades e até cultivavam seguidores e fãs” (Rea, 2016, p. 23,

tradução nossa). Isso atraiu ainda mais patrocínios e clientes às PC bangs, e esse período pode ser considerado o início dos esportes eletrônicos profissionais na Coreia do Sul.

Com a fundação de um órgão regulador dos *esports* em 2000, a *Korea Esports Association* (KeSPA) trouxe relevâncias maiores ao meio, com a inclusão de patrocínios de grandes empresas, como Samsung, Korean Air e Asiana nos torneios (Yu, 2015). A expansão também significou prêmios em dinheiro cada vez maiores e competições mais disputadas entre jogadores. Shin Ju-young, o primeiro jogador coreano a ser registrado na KPGL, alcançava horas de jogo que eram em torno de 15 horas por dia (Jin, 2020).

Conforme demonstrado anteriormente, o impacto do rápido crescimento do meio teve repercussões econômicas e socioculturais significativas na Coreia do Sul. No ano 2000, o cargo de *gamer professional* foi reconhecido como uma categoria oficial de emprego pelo governo coreano (Rea, 2016; Esports Insider, 2023). Em pesquisa realizada em 2001 pelo Instituto de Pesquisa da Coreia de Educação e Treinamento Vocacionais, o cargo *gamer professional* assumiu pela primeira vez a posição de profissão mais popular escolhida entre alunos do ensino fundamental (Jin, 2020). Em 2004, a final da liga profissional de StarCraft atraiu mais de 100.000 fãs à praia de Busan para prestigiar o evento (Mozur, 2014).

2.4. O surgimento do *League of Legends*

É lançado em 2002 o jogo *Warcraft III: Reign of Chaos*, publicado pela mesma desenvolvedora do grande sucesso StarCraft, a Blizzard. Seguindo a mesma temática de jogo de estratégia em tempo real (RTS), o jogo atraiu tanto fãs antigos da franquia Warcraft, quanto novos jogadores que desenvolveram interesse pela categoria RTS em anos recentes. Apesar disso, a principal influência do jogo para a história dos *esports* se deu pela possibilidade da inclusão de *mods* – modificações feitas ao jogo criadas por jogadores.

Um "mod" em um jogo, abreviação de "modificação" (do inglês "*modification*"), refere-se a uma alteração ou personalização feita por jogadores em um jogo existente para adicionar, alterar ou expandir seus recursos, mecânicas ou conteúdo. Os mods são criados por fãs da comunidade de jogadores e podem variar amplamente em termos de complexidade e propósito.

Em 2005, o mod *Defense of the Ancients*, comumente conhecido como DOTA foi lançado e alcançou grande popularidade entre a comunidade do jogo. O DOTA introduziu a ideia de equipes de jogadores controlando heróis únicos, cada um com suas próprias habilidades e papéis no jogo. Isso incentivou uma jogabilidade altamente cooperativa e competitiva, na qual as equipes se enfrentavam em uma arena com o objetivo de destruir a base inimiga. A popularidade crescente do DOTA levou ao desenvolvimento de várias versões aprimoradas do jogo e à

criação de um gênero derivado do RTS completamente novo conhecido como MOBA (*Multiplayer Online Battle Arena*).

Em 2006, dois entusiastas do DOTA e colegas de quarto, Brandon Beck e Marc Merrill, fundaram a Riot Games com o objetivo de “criar jogos que pudessem ser aproveitados por jogadores ávidos como eles, o tipo de jogos que eles poderiam passar centenas de horas jogando, testando suas habilidades.” (Founding... 2016).

A fim de lançar um jogo que pudesse superar o DOTA, Brandon e Marc iniciaram os trabalhos para o lançamento do primeiro jogo da Riot Games: o *League of Legends* (LoL). Com otimizações de qualidade de vida e sistemas dedicados inteiramente ao estilo MOBA, os criadores esperavam que o LoL trouxesse mecânicas únicas e sistemas mais competitivos que recompensassem jogadores de maior habilidade. Além disso, os criadores do jogo também tinham o objetivo de publicá-lo de forma gratuita, de tal forma que o jogo pudesse alcançar uma comunidade ainda maior.

Após o lançamento do jogo em 2009, o LoL tornou-se um fenômeno de escala global e até a data atual de setembro de 2023 a Riot Games sediou 2893 torneios com um total acumulado em premiações de US\$ 104.914.716,06 e maior premiação em torneio único de US\$ 6.450.000,00, referente ao campeonato mundial de 2018 (Esports Earnings, 2023).

2.5. O sucesso mundial da Coreia do Sul

O primeiro torneio mundial de *League of Legends* teve início em 2011, em *Jönköping*, na Suécia. Com a premiação total de 100 mil dólares americanos, o torneio contava com 8 times, sendo estes: 3 da América do Norte, 3 da Europa e 2 do Sudeste Asiático. Com a vitória do time europeu Fnatic, o primeiro torneio mundial foi o único que contou com um time vencedor ocidental até o momento, em 2023. O torneio teve um pico de espectadores simultâneos de 210 mil pessoas (Mitchell, 2011).

O segundo torneio mundial, sediado nos Estados Unidos em 2012, contou pela primeira vez com times coreanos e chineses após o lançamento de servidores oficiais em mais países. Já de início, os times coreanos eram cotados como favoritos para vencer (Almughrabi, 2021). Apesar disso, o time coreano alcançou apenas o segundo lugar, perdendo para a equipe taiwanesa Taipei Assassins. A final atraiu mais de 1.1 milhão de telespectadores simultâneos, tornando-se, até o momento da época, o evento de *esports* mais assistido de todos os tempos (Goetomo, 2016, p. 13, apud Breslau, 2012).

Embora um time sul-coreano não tenha ganhado a competição em primeira representação do país no mundial, a diferença de investimento e preparação dos times coreanos

em relação às outras regiões se mostrou por meio de vitórias dominantes durante a fase de grupos e chave eliminatória, com ambos os times coreanos perdendo apenas para o time taiwanês vencedor, e uma derrota de um jogo em uma série melhor de três para a CLG.EU, um time europeu. Enquanto outras regiões começavam a explorar o meio dos *esports*, a Coreia já trazia em sua cultura da informação a transmissão de partidas no sinal de televisão aberto, entrevistas e documentários de jogadores, e no geral uma infraestrutura para *esports* que podia ser facilmente comparada com àquelas destinadas a atletas de esportes tradicionais (Yu, 2015).

Em 2013, com o objetivo de expandir o formato do campeonato mundial para incluir as novas regiões e permitir competições locais, as ligas competitivas regionais foram lançadas, com destaques para as *League Championship Series (LCS)* da América do Norte (NA LCS) e da Europa (EU LCS), que contavam com grande cobertura midiática e o começo da profissionalização do *gamer* no ocidente.

Também em 2013, começava a história de uma das maiores dinastias dos *esports*. O time coreano SK Telecom T1, sob o comando de seu jogador estrela, Lee Sang-hyeok, comumente conhecido “Faker” alcançava o primeiro de três campeonatos mundiais a serem conquistados nos próximos anos. A final foi contra o time chinês Royal Club, e o evento contou com 32 milhões de espectadores únicos, e um pico de 8.5 milhões de espectadores simultâneos (McCormick, 2013), valores acima de oito vezes maiores do que o ano anterior. Como forma de comparação, o número de espectadores foi maior, em 5 milhões, do que a audiência obtida pelas finais da liga de basquete norte-americana, a NBA (Schwartz, 2014).

A Coreia do Sul voltou a vencer pelos próximos 4 anos seguidos em o que só pode ser descrito como uma dominação completa da região sob qualquer outro tipo de competição. As finais de 2015, 2016 e 2017 foram disputadas entre dois times coreanos, e a região só começou a perder a sua hegemonia por nova competição de times chineses, que assumiam um estilo de jogo mais agressivo e destoavam do padrão de outros times de querer copiar o estilo de jogo coreano para alcançar o sucesso. É durante esse período também que muitos jogadores e profissionais sul-coreanos da área começaram a ser contratados por times de outros países, com destaque para a rápida inclusão de coreanos no cenário competitivo brasileiro. Em 2014, dos 34 jogadores registrados na liga brasileira, havia 4 sul-coreanos divididos entre dois times na competição – times estes, que chegaram à final do campeonato e demonstraram dominação sob oponentes na fase eliminatória da competição, não perdendo nenhuma partida.

Ano após ano o campeonato mundial quebrava recordes de audiência e a notoriedade da competição aumentava. Em 2015, o número de horas assistidas quase dobrou o valor de 194 milhões do ano anterior, alcançando a marca de 360 milhões de horas. O evento também atraiu

número recorde de espectadores únicos, com 36 milhões, e espectadores simultâneos, com 14 milhões (Figueira, 2015). Em 2016, 43 milhões de espectadores únicos (Howell, 2016), e em 2017, 80 milhões (Goslin, 2017).

Com 2018 sendo o primeiro ano em que nenhum time sul-coreano alcançou a final devido a uma série de partidas inesperadas e o desenvolvimento do jogo em outras regiões, o evento também teve seu recorde de público, alcançando 99.6 milhões de espectadores únicos durante a competição (Lol esports Staff, 2018).

Desde então, o LoL se solidificou como um dos mais populares jogos do mundo, e contou em 2022 com 180 milhões de jogadores registrados, com mais de 32 milhões de jogadores ativos em um único dia (Kamberovic, 2023). Em relação aos campeonatos mundiais, times sul-coreanos venceram em 7 dos 12 campeonatos existentes, e a China assumiu uma nova colocação como uma das favoritas nos torneios, vencendo os campeonatos de 2018, 2019 e 2021. Apesar disso, o servidor local coreano ainda é tido como um dos de maior prestígio do jogo, e muitos jogadores profissionais e fãs viajam para o país para poder jogar no servidor com os profissionais coreanos – viagens como essas para poder jogar em outros servidores são popularmente conhecidas como *bootcamps*, e se tornaram populares em períodos anteriores ao campeonato mundial para que os jogadores possam estar em sua melhor forma para a competição.

A Riot Games atualmente gere 11 servidores, mas conta com mais 19 servidores geridos pela China Telecom e 7 pela China Netcom na República Popular da China e um servidor do Sudeste Asiático, gerido pela Garena, um parceiro da Riot Games (Runas LOL, 2021). Os servidores chineses, apesar de não serem hospedados diretamente pela Riot, são também de grande influência nos *esports* e, similarmente à Coreia, possuem muitas vezes ambientes de jogo fechados e inacessíveis devido às limitações de acesso do governo, ou pela barreira linguística.

O servidor coreano é também notavelmente o único servidor gerido pela Riot Games que exige um número de documento oficial do governo— seria, no caso do Brasil, o número do CPF ou RG, por exemplo – do jogador para a criação da conta. Isso torna o servidor um dos mais exclusivos do mundo e muitas vezes inacessível ao público geral. Um dos possíveis motivos para essa limitação é a quantidade de jogadores de países vizinhos que criavam contas no servidor coreano por ele ser um dos primeiros a ser lançado no continente asiático em 2011. O servidor japonês, por exemplo, só teve o seu lançamento em 2016, então muitos jogadores japoneses compravam contas de jogadores coreanos ou burlavam cadastros para criar uma conta no servidor sul-coreano.

Tabela 1 – Servidores geridos pela Riot Games

SERVIDOR	SIGLA	DATA DE LANÇAMENTO
América Do Norte	NA	2009
Europa Ocidental	EUW	2010
Europa Nórdica & Leste	EUNE	2010
República Da Coreia	KR	2011
Brasil	BR	2012
Turquia	TR	2012
América Latina Norte	LAN	2013
América Latina Sul	LAS	2013
Oceania	OCE	2013
Rússia	RU	2013
Japão	JP	2016
Ambiente Público Beta	PBE	2009

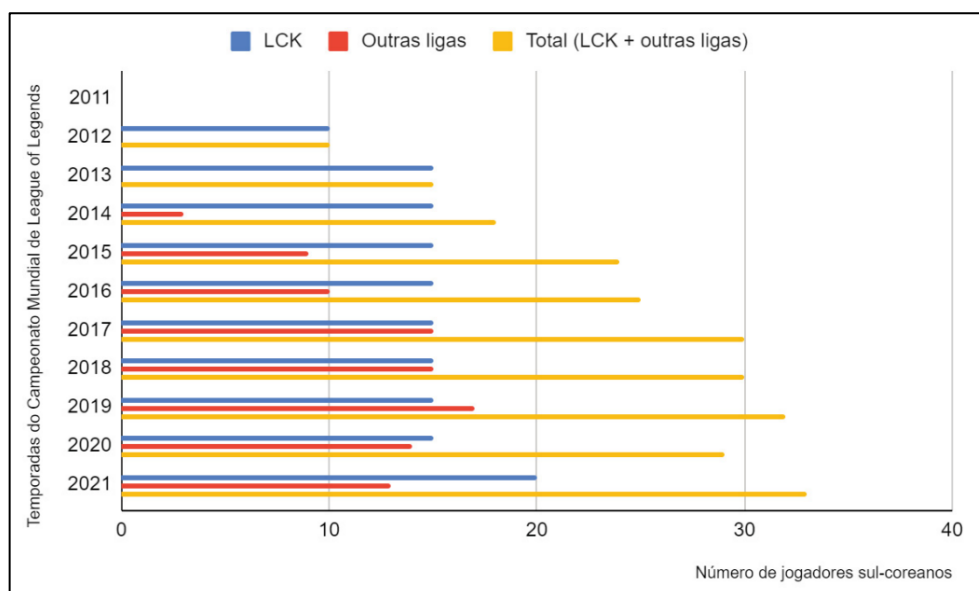
Fonte: RUNAS LOL, 2021

Enquanto a população da Coreia do Sul é composta por 51,8 milhões de pessoas (UNPF, 2023), o servidor coreano de League of Legends é o maior servidor do mundo em número de jogadores, com um número de contas ativas de 3,392 milhões de contas (OP.GG, 2023), o que, em números brutos e descartando fatores como contas criadas por estrangeiros e contas duplicadas, representaria aproximadamente 15% de toda a população sul-coreana. Entende-se por contas ativas aquelas que alcançaram colocações no sistema ranqueado de 2023, o que simboliza a realização de ao menos dez partidas no sistema ranqueado do jogo. Considerando a distribuição demográfica coreana e a cultura de PC Bangs pode se inferir que, efetivamente, o número em percentual de jogadores ativos de idades de 18-30 anos seja ainda maior do que valores para a população coreana no geral. Além disso, League of Legends é o jogo mais jogado do país em PC Bangs, ocupando cerca de 40% de toda a distribuição dos jogos jogados, seguido pelo FIFA Online, com 10% (Gametrics, 2023). Em comparação, o Brasil, mesmo com uma população mais do que quatro vezes maior do que o país sul-coreano, possui apenas 641.232 contas ativas no servidor brasileiro (OP.GG, 2023). O número do servidor brasileiro é ainda mais chocante em comparação ao sul-coreano ao levar em consideração que a população brasileira que joga algum tipo de jogo eletrônico corresponde a 74,5% da população total (Pesquisa Game Brasil, 2022).

Ao analisarmos o número de profissionais coreanos que atuam no exterior em ligas oficiais do League of Legends, podemos perceber mais claramente a influência sul-coreana. Conforme banco de dados *League of Legends Esports League-Recognized Contract Database* para o ano de 2023 (RIOT GAMES, 2023), na liga norte-americana de League of Legends, a League Championship Series (LCS), dos 93 profissionais registrados, 17 são de nacionalidade sul-coreana. No Circuito Brasileiro de League of Legends (CBLLOL), 14 dos 73 profissionais registrados são de nacionalidade sul-coreana. A evolução da participação coreana em ligas internacionais também pode ser vista na participação de jogadores sul-coreanos nos campeonatos mundiais em times de outras regiões.

O gráfico elaborado por Ferreira (2022) demonstra uma participação crescente de jogadores sul-coreanos no Campeonato Mundial por meio de outras ligas além da coreana. Em 2017 e 2018, havia a mesma quantidade de jogadores sul-coreanos divididos entre times da liga coreana e de outras ligas. Em 2019, o número de jogadores sul-coreanos presentes em ligas internacionais foi maior do que a de jogadores em times da liga coreana. É interessante perceber também que, em 2021, os jogadores sul-coreanos representaram quase 40% do total de jogadores inscritos na competição, com 32 dos 82 jogadores inscritos sendo de nacionalidade sul-coreana:

Gráfico 1 – Número de jogadores sul-coreanos participantes do Campeonato Mundial por liga



Fonte: Ferreira (2022)

Com a inclusão de jogadores coreanos na liga brasileira há 9 anos atrás – 2014, a metodologia de trabalho coreana e influência cultural no geral foram algumas das características incluídas no cenário competitivo brasileiro. Porém, apesar de haver um consenso geral da

comunidade de que a inclusão de estrangeiros no campeonato brasileiro tenha elevado o nível de jogo da região, o Brasil é uma das regiões com um dos piores históricos de desempenho em campeonatos mundiais – desde 2014 até o campeonato mundial de 2022, a região conta com 16 vitórias e 41 derrotas em partidas oficiais pelo torneio, uma taxa de vitórias inferior a 40%. Além disso, por ser considerado uma região *minor*, de nível inferior, apenas o vencedor da liga brasileira se classifica para o mundial; ou seja, enquanto há outras regiões *major* como a Coreia e a China com mais vagas para a competição, a chance de representação brasileira se resume apenas a um time.

É controverso simplesmente titular o vencedor de uma competição como o melhor time dada a imprevisibilidade do esporte e inúmeras possibilidades de resultados em jogos. Porém, se fossemos assumir o time representante brasileiro no campeonato mundial como o melhor time da região por ter ganhado o campeonato, outro dado interessante é descoberto. Desde 2014, em que houve a primeira inclusão de jogadores sul-coreanos na região, até 2022, apenas 2 dos 9 times que representaram a região brasileira possuíam sul-coreanos em seu elenco de jogadores. De antemão, para a região norte-americana, o time com a melhor colocação classificado para o mundial – o time *seed 1* – possuía jogadores sul-coreanos em 6 das 9 ocasiões.

Apesar do sucesso coreano nos *esports* se demonstrar inúmeras vezes ao passar dos anos, percebe-se que é incerto afirmar que o mesmo sucesso foi demonstrado em meio ao cenário brasileiro, uma vez que, na maioria das vezes, o campeonato foi ganho por equipes compostas inteiramente por brasileiros. Com um jogo em constante evolução e seguidos anos de fracasso da região em conseguir mais vitórias internacionais, o questionamento do porquê essa dominação de coreanos não funcionar claramente no Brasil é um tema recorrente na comunidade de jogadores, e apesar de não haver uma resposta simples, dois pontos principais são centrais ao tema: a barreira linguística e as diferenças culturais.

2.6. Limitações para a difusão do conteúdo coreano a outras comunidades

A barreira linguística é um grande desafio nos *esports*, especialmente quando se trata da integração de jogadores estrangeiros em equipes brasileiras. Em competições de alto rendimento, a comunicação precisa ser precisa e eficaz, e a falta de fluidez na comunicação entre jogadores brasileiros e coreanos pode afetar o desempenho do time, uma vez que ambos são obrigados a utilizar o inglês como língua neutra entre o português e o coreano.

Por mais que o inglês seja frequentemente considerado uma língua comum para a comunicação internacional nos *esports*, é importante observar que o Brasil tem uma relação complexa com o idioma. De acordo com a pesquisa do British Council (2013), apenas 5,1% da

população brasileira afirmou ter algum conhecimento da língua inglesa. Além disso, entre esse grupo, apenas 16% se consideram fluente no idioma. Isso significa que a vasta maioria dos brasileiros não possui habilidades avançadas em inglês, tornando-o menos eficaz como meio de comunicação em um ambiente de alto rendimento.

Além da barreira linguística, há diferenças culturais altamente impactantes. A Coreia do Sul possui uma sociedade altamente competitiva. Jin (2020) demonstra que a Coreia é uma sociedade movida pela competição:

“Embora os atletas olímpicos se tornassem heróis nacionais na maioria dos países quando recebessem alguma medalha, um atleta coreano que ganhasse uma medalha de prata nas Olimpíadas muitas vezes abaixaria a cabeça de vergonha e pediria desculpas ao público por não ter alcançado o ouro. (Jin, 2020, p. 3739 tradução nossa)”.

Na Coreia, o desempenho é altamente valorizado e a pressão para alcançar a excelência é constante. A mentalidade de perseguir o primeiro lugar é uma norma cultural do país, e isso se estende aos *esports*, o que pode entrar em conflito com a cultura brasileira em relação à filosofia de treinos e de trabalho em equipe. Muitos jogadores sul-coreanos que vieram ao Brasil retornaram ao seu país após períodos inferiores a 1 ano, demonstrando dificuldades de adaptação à cultura e à língua.

Embora haja exemplos de jogadores coreanos que conseguiram se adaptar ao cenário brasileiro e alcançar o sucesso regional, esses desafios culturais e linguísticos continuam a ser obstáculos a serem superados. À medida que o cenário de *esports* brasileiro continua a evoluir e amadurecer, encontrar maneiras de integrar jogadores estrangeiros de forma mais eficaz e facilitar a comunicação entre equipes multiculturais pode ser crucial para o sucesso futuro da região nos campeonatos mundiais de League of Legends e outros jogos competitivos.

Enquanto a indústria de *esports* coreana já é consolidada e as equipes são partes de empresas autossuficientes sob a perspectiva econômica, a indústria brasileira ainda é relativamente nova e apenas recentemente começou a receber patrocínios de empresas mais consolidadas e ter divisões de departamentos dentro da empresa. Muitas vezes pela falta de recursos, ainda são raros os tradutores e profissionais com cargos dedicados a facilitar essa integração multicultural nos times.

Apesar do fracasso internacional, o público brasileiro nos *esports* é considerado um dos mais apaixonados e investidos no jogo. Em 2021, o pico de espectadores do CBLOL foi inferior somente às ligas europeia e coreana, com um total de 416.335 espectadores simultâneos

(Esports Charts, sem data). O desenvolvimento da indústria local e um melhor desempenho internacional seriam fatores essenciais para o crescimento dos *esports* no Brasil.

De tal forma, propõe-se a criação de um corpus paralelo entre o coreano e o português de termos dos *esports*, com enfoque em League of Legends devido à sua relevância, com a finalidade de diminuir a barreira linguística atual entre os países. Tal corpus serviria como uma ponte linguística entre duas línguas tão distintas, permitindo que o vasto conteúdo produzido em coreano no contexto de *esports* seja acessível aos falantes de português.

A Coreia do Sul, como um dos epicentros globais dos *esports*, produz uma quantidade significativa de materiais, como vídeos de treinamento, análises de partidas, guias estratégicos e entrevistas com jogadores e treinadores. No entanto, grande parte desse conteúdo permanece inacessível para o público brasileiro devido à falta de tradução e à barreira linguística. Isso cria uma lacuna no conhecimento disponível para os jogadores e fãs de *esports* no Brasil, que ficam fora da “bolha” de conteúdos de alta qualidade coreanos.

A criação de um corpus paralelo entre o coreano e o português não apenas facilitaria o acesso ao conteúdo coreano para o público brasileiro, mas também contribuiria para o desenvolvimento do cenário de *esports* no Brasil. Com uma compreensão mais abrangente e aprofundada das práticas e estratégias adotadas pelos jogadores sul-coreanos, as equipes brasileiras podem se beneficiar ao incorporar elementos bem-sucedidos em suas próprias estratégias de jogo.

Além disso, espera-se que a disponibilidade de um corpus paralelo sirva como incentivo para o intercâmbio cultural entre as duas comunidades de *esports*. Isso pode levar a uma maior colaboração entre jogadores e treinadores brasileiros e sul-coreanos, proporcionando oportunidades de aprendizado e crescimento mútuo. À medida que o cenário competitivo de *esports* continua a se expandir globalmente, a capacidade de se comunicar e compartilhar conhecimento entre diferentes regiões se torna cada vez mais crucial. De tal forma, a linguística de corpus torna-se um agente relevante em meio à expansão de conteúdos relevantes nos *esports*.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. Linguística de Corpus

Para a elaboração de um corpus, é necessário primeiro compreender o que é um corpus e quais são as áreas de aplicação da Linguística de Corpus. A Linguística de Corpus computacional é uma área da ciência relativamente recente, com o lançamento do primeiro

corpus linguístico eletrônico apenas em 1964: o corpus Brown. Composto de 1 milhão de palavras, o corpus ainda é considerado um dos maiores pioneiros de todos os tempos pela quantidade de dados adquirida na época, uma vez que os textos eram transferidos por cartões físicos, perfurados a mão, para o computador (Sardinha, 2000). Desde então, a Linguística de Corpus expandiu-se tanto em tamanho dos corpora – hoje é possível encontrar corpora com mais de um bilhão de tokens – quanto em áreas de aplicação, abrangendo perspectivas didáticas, acadêmicas, comerciais e de processamento linguístico. Tony Sardinha também traz em sua obra “Linguística de Corpus” uma breve descrição do tema e de sua aplicação:

A Linguística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (Sardinha, 2004, p. 3)

Um corpus é uma coleção estruturada de textos ou dados linguísticos, geralmente reunidos para fins de análise linguística, pesquisa e estudo. Essa coleção pode abranger uma variedade de gêneros textuais e serve como uma representação sistemática da linguagem em seu contexto de uso. Sardinha (2000) define corpus como: 'corpo', 'conjunto de documentos', e enfatiza a existência de corpora antes da criação dos computadores, com exemplos do Corpus Helenístico na Grécia Antiga, e corpora de citações da Bíblia na Idade Média.

3.2. Corpus paralelo

Em nosso estudo, busca-se a criação de um tipo específico de corpus: o corpus paralelo. Estamos focados em analisar e traduzir textos relacionados aos *esports* coreanos; por exemplo, se quisermos entender como um termo específico relacionado aos *esports* é utilizado em coreano e encontrar sua tradução adequada para o português, um corpus paralelo nos permite fazer isso de maneira eficaz.

Um corpus paralelo, também conhecido como corpus bilíngue – mas podendo também ser multilíngue, como é o caso do corpus que iremos construir – é uma coleção de textos ou dados linguísticos em duas ou mais línguas que estão alinhados de forma que seja possível identificar correspondências entre segmentos linguísticos equivalentes. Ou seja, é uma ferramenta que permite a análise de textos em diferentes idiomas em paralelo, possibilitando a identificação de traduções e equivalências linguísticas (Baker, 1993).

Diferentemente de corpus como o COMPARA, composto por textos originais do inglês traduzidos para o português e de textos originais do português traduzidos para o inglês, ou seja,

um corpus paralelo bidirecional (Dayrell, 2015, p. 93), o corpus que elaboraremos é unidirecional. Devido à falta de traduções de reputação confiável do tema de textos originários do português para o coreano ou do inglês para o coreano, a proposta inicial visa apenas trazer textos originais do coreano para o português e o inglês.

A possibilidade de utilização de um corpus paralelo é diretamente ligada a disponibilidade de textos e de traduções. Por esse motivo, textos oficiais da União Europeia são frequentemente utilizados nessa modalidade (Olohan, 2004). A escassez de recursos bilíngues coreano-português fogem da aplicação comum dos corpus paralelo – mas é ao mesmo tempo o maior motivador para a elaboração do corpus.

3.3. Estudos de tradução

Johansson (2007) afirma que por meio do corpora multilíngue podemos ver como línguas se assimilam e se diferem ao compará-las em escala maior, podendo possivelmente encontrar características sobre a língua geral como forma de comunicação humana. Porém, é também por meio do corpora que podemos realizar inúmeras avaliações e trabalhar com os textos por meio de diferentes abordagens.

A principal abordagem retratada pelo autor é a da análise contrastiva (AC), que é a “comparação sistemática de duas ou mais línguas, com o objetivo de descrever suas semelhanças e diferenças.” (Johansson, 2007, tradução nossa). Muitas vezes utilizada no meio acadêmico para o ensino de línguas, tal abordagem traz problemas quando há uma falta de equivalência direta, e por isso necessita de estudos tradutórios e estudos comparativos para ter materiais suficientes.

Os estudos tradutórios incluem uma série de abordagens, desde pesquisas teóricas e descritivas sobre fenômenos de tradução até questões aplicadas, como avaliação de tradução e treinamento de tradutores. Por um lado, é, portanto, mais amplo do que a análise contrastiva. Noutro, é mais restrito, pois limita-se aos textos que são traduzidos e às formas como são transmitidos entre línguas. (Johansson, 2007, p. 4, tradução nossa)

Uma vez que estudos tradutórios vão além da comparação sistemática entre línguas, mas ao mesmo tempo são limitados a disponibilidade das traduções, é necessário compreender a sua bidirecionalidade com a AC, de forma que a tradução fornece dados para a análise contrastiva, e a análise auxilia a tradução a solucionar problemas encontrados no processo (Hoey; Houghton, 1998, apud Johansson, 2007). Para alcançar os objetivos adequados de

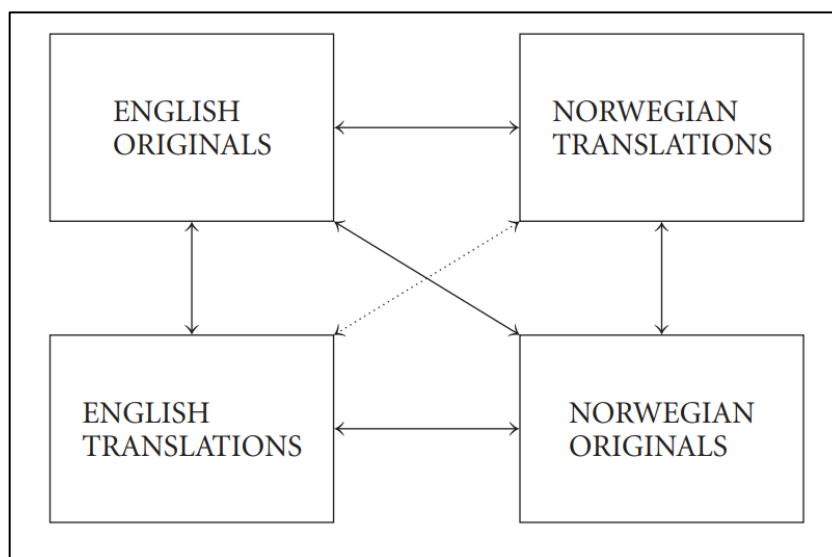
análise linguística após a criação de um corpus, o modelo e a formatação do corpus paralelo devem ser bem estabelecidos, possibilitando o diálogo entre estudos tradutórios e comparativos.

Entendem-se corpora que seguem o modelo tradutório os que contém textos originais com traduções para uma ou mais línguas, e corpora comparativos os que contém textos originais em duas ou mais línguas interligadas por critérios como gênero, período de publicação e outras características (Johansson, 2007).

3.4. Estruturação do corpus

Tomando como base o modelo do Corpus Paralelo Inglês-Norueguês, em inglês, *English-Norwegian Parallel Corpus (ENPC)*, a estruturação do nosso corpus será feita de maneira similar ao esboçado por Johansson (2007) para o ENPC, corpus tomado como exemplo de corpus tradutório e comparável em sua obra. As maiores diferenças são a ausência de bidirecionalidade de textos do inglês e português para o coreano devido à falta de recursos, e a inclusão de um terceiro idioma, o que torna o nosso corpus multilíngue.

Figura 1 – Estruturação do *English-Norwegian Parallel Corpus (ENPC)*



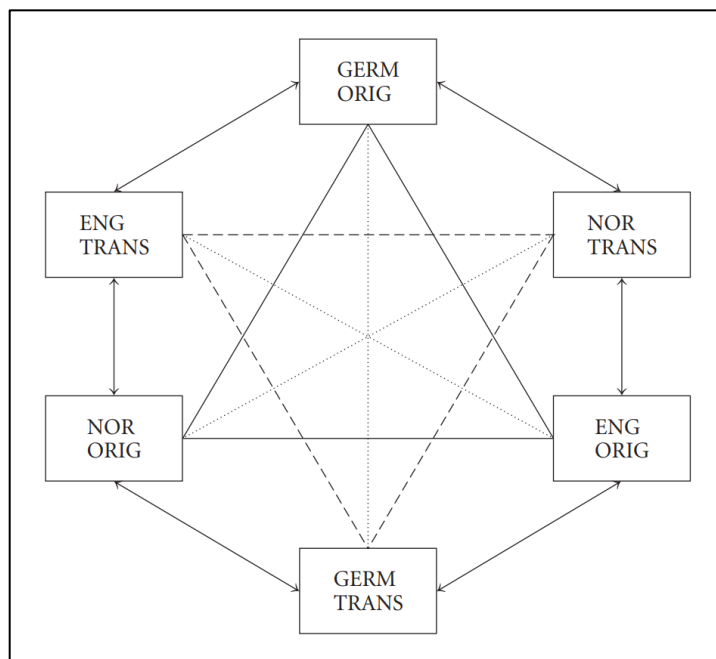
Fonte: Johansson (2007)

O corpus ENPC assume múltiplas funções. Pode ser utilizado como um corpus de tradução ao comparar textos originais e suas traduções entre os idiomas (setas horizontais); como corpus comparável entre textos originais (seta diagonal sólida e setas verticais) e entre textos traduzidos (seta diagonal pontilhada) (Johansson, 2007).

Uma extensão do corpus ENPC, o *Oslo Multilingual Corpus (OMC)* segue um modelo alternativo que é mais viável para a criação de um corpus plurilíngue – no caso específico desse

corpus, o inglês, o alemão e o norueguês. O modelo, denominado de “modelo estrela” por Johansson (2007) pode ser visto na Figura 2:

Figura 2 – Estruturação do *Oslo Multilingual Corpus (OMC)*



Fonte: Johansson (2007)

A razão pela qual o modelo não pode ser seguido de forma idêntica é a falta de textos traduzidos de outras línguas para o coreano, o que impediria um dos pilares dessa interação entre três línguas. Tal limitação gerada pela falta de bidirecionalidade em línguas específicas é demonstrada em prática na comparação entre um possível corpus com base na língua inglesa e outro com base na língua norueguesa:

Para uma língua como o inglês, o modelo estrela poderia facilmente ser estendido a muitas línguas, visto que as traduções são abundantes. Para o norueguês como língua de origem, há muito menos traduções e é difícil encontrar textos que sejam traduzidos para várias línguas, uma vez que as preferências pelos textos selecionados para tradução diferem consideravelmente dependendo da comunidade linguística de destino. (Johansson, 2007, p. 19, tradução nossa)

Considerando as limitações da língua norueguesa similares às da língua coreana sob essa perspectiva de falta de recursos, o presente corpus busca solucionar o problema por meio de documentações oficiais dos jogos eletrônicos, uma vez que possuem adaptação e localização de recursos entre vários idiomas. É inegável que a seleção de tais textos não necessariamente demonstram o vocabulário utilizado pela comunidade de jogadores e falantes da língua alvo, porém foi selecionada devido a maior facilidade de adaptação ao modelo.

3.5. Terminologia – TGT e TCT

Com o objetivo de gerar um corpus que permita melhor entender termos utilizados no domínio dos *esports*, é essencial que se entenda um pouco sobre a Terminologia para realizar a sua identificação e classificação de maneira precisa. A fim de facilitar o entendimento, a palavra Terminologia será grafada com o “T” maiúsculo ao ser referida como um campo de estudos ou disciplina, e com o “t” minúsculo quando referida com o sentido de conjunto de termos técnico científicos (Krieger e Finatto, 2004, p. 13). Tal abordagem terminológica é essencial ao compreender o domínio dos *esports* como uma área de especialidade.

Apesar da terminologia em si ser uma prática antiga, uma vez que os saberes técnicos e científicos carregam com si uma língua de especialidade desde suas criações, a Terminologia como campo de estudo emergiu somente no século XX com Eugen Wüster, engenheiro austríaco considerado o fundador da terminologia moderna (Cabré, 1993; Krieger; Finatto, 2004). Wüster introduziu a Terminologia como disciplina na Universidade de Viena em 1972. Segundo Krieger e Finatto (2004, p. 20), Wüster tinha o objetivo de “padronizar o uso de termos técnicos-científicos de modo a alcançar a univocidade comunicacional no plano internacional”. Isso resultou em uma série de estudos que levaram ao desenvolvimento da TGT, a Teoria Geral da Terminologia.

A TGT está fortemente vinculada “ao objetivo de delinear diretrizes pragmáticas de normatizar as terminologias, visando a facilitar seu uso unívoco mundialmente.” (Krieger; Finatto, 2004, p. 28). Esforços realizados por Wüster para garantir a univocidade da terminologia podem ser vistos em aplicação por meio de comitês internacionais que garantem a manutenção da terminologia e normas técnicas, como o Comitê 37 da ISO (*International Standard Organization*), *Office de la Langue Française*, no Quebec, e TERMCAT, Centro de Terminologia da Catalunha, em Catalunha.

Apesar de garantir maior padronização do léxico de especialização a fim de garantir univocidade terminológica, a abordagem clássica da TGT de Wüster considerava os termos como unidades cognitivas, excluindo aspectos lexicais, como diacronia, sinonímia e variação. Essa visão foi revista à medida que a Terminologia se desenvolveu. Novas abordagens socioterminológicas, comunicativas, cognitivas, éticas e textuais-discursivas surgiram para considerar o termo não apenas como uma entidade cognitiva, mas também como uma unidade lexical que opera em contextos de uso. A Teoria Comunicativa da Terminologia (TCT), desenvolvida por Maria Teresa Cabré contesta a TGT, especialmente por ser uma teórica descritiva, e não normativa. A TCT enfatiza o contexto, a interação linguística e a flexibilidade

do termo – como variações conceituais e denominativas (polissemia e sinonímia), permitindo-lhe adaptar-se a diferentes contextos comunicativos.

Ao avaliar as duas teorias apresentadas, TGT e TCT, o presente trabalho se alinha melhor à Teoria Comunicativa da Terminologia (TCT) devido à sua natureza descritiva e expansão de definição única dos termos. Acredita-se que a observação de uso dos termos em seu ambiente de ocorrência possibilita perceber variações conceituais e denominativas a partir de seu contexto de uso, uma vez que termos podem ser polissêmicos a depender do ambiente de uso.

3.6. Terminologia no corpus

O processo de criação do corpus influencia diretamente em sua interação com a Terminologia, conforme visto por Krieger e Finatto:

[...]quando se busca se identificar termos em textos com o apoio do computador, depende-se de fatores tais como o tamanho e o tipo de um *corpus* textual utilizado como fonte. O *corpus* deve ser definido de acordo com as especificidades e objetivos de cada trabalho e deve ser minimamente representativo. (Krieger; Finatto, 2004, p. 205)

Dessa forma, é interessante revisitar orientações demonstradas na obra de Almeida e Correia, *Terminologia e corpus: relações, métodos e recursos* (2008), como recurso-base para a criação de um corpus que permita uma pesquisa terminológica de maior qualidade. Em linha similar à Krieger e Finatto (2004), Almeida e Correia (2008) também defendem a importância de uma elaboração bem-conceituada do corpus, uma vez que o corpus compilado interfere diretamente com todas as etapas posteriores.

Sendo assim, a primeira informação relevante é elaborar um corpus em que seja possível realizar a extração de candidatos a termos, ou seja, “itens léxicos que se comportam nos seus respectivos contextos como termos, mas cuja autenticidade será validada posteriormente” (Almeida; Correia, 2008). É importante notar, em especial, duas características de tal afirmação: de que o contexto tem a sua devida importância, conforme imaginado pela TCT, que terá devida preferência em análise terminológica, e de que há a necessidade de uma validação de autenticidade terminológica.

A validação pode ser feita de diferentes formas, com uma recomendação maior àquela realizada de forma manual por especialistas do domínio. Apesar de ser o método recomendado, por limitação de tempo e disponibilidade de profissionais capazes, esta alternativa é muitas

vezes substituída uma comparação da lista de candidatos a um corpus de referência e contextualização de uso a partir dessas informações (Almeida; Correia, 2008, p. 75).

Referente à compilação do corpus, Almeida e Correia apontam seis etapas a serem seguidas para possibilitar pesquisas terminológicas:

A primeira etapa na criação de um corpus terminológico é a "delimitação do domínio". Esta etapa visa restringir a área de estudo, evitando abordagens excessivamente abrangentes que podem se mostrar ineficazes. Ao delimitar um domínio, o pesquisador é capaz de secionar subáreas de especialidade com contexto específico, o que impede diferenças significativas entre subáreas, ou necessidade de um trabalho extenso e complexo em um único corpus – tornando mais aconselhável a sua divisão em corpora para facilidade de análise após sua completude. (Almeida; Correia, 2008, p. 77-79).

A segunda etapa é a "seleção das fontes de onde provirão os textos que deverão compor o *corpus*, bem como atenção aos requisitos recomendados e que já são consenso na Linguística de *Corpus*". Nessa etapa, as autoras reforçam a necessidade de escolher fontes confiáveis para a construção do corpus. Essas fontes podem ser identificadas por meio de instituições, entidades, universidades e laboratórios envolvidos no domínio em questão, com preferência por fontes que disponham de material escrito, especialmente em formato digital. Também deverão ser levantadas questões como quantidade de termos, formatação do produto terminológico a ser gerado, público-alvo, variedade de fontes (riqueza textual) e outros fatores relevantes. A natureza do corpus a ser criado influenciará e será influenciado pela seleção dos textos que o comporão. (Almeida; Correia, 2008, p. 79-82).

A terceira etapa, denominada "compilação (ou captura) dos textos", envolve a coleta e o armazenamento dos textos escolhidos em arquivos. Essa etapa pode ter grande variação a depender da fonte utilizada para a compilação do corpus, variando de uma simples cópia de arquivos de texto à necessidade de realizar um reconhecimento óptico de caracteres (OCR) para realizar a sua extração. Há vários métodos e casos de uso específicos para essa etapa. (Almeida; Correia, 2008, p. 82-84).

A quarta etapa, "conversão de formatos e limpeza dos textos", requer a conversão, manual (recurso "copia-e-cola") ou automática, por meio de softwares disponíveis para a conversão de arquivos, como o XPDF, citado pelas autoras como alternativa gratuita. É recomendada a conversão de todos os arquivos para o formato "txt", dada a sua facilidade de manipulação. Nesse estágio, os textos também passam por um processo de limpeza e formatação, removendo elementos que fogem do formato textual, como imagens, gráficos e tabelas. (Almeida; Correia, 2008, p. 84).

Apesar da forte recomendação pelo uso do formato “txt” na conversão dos arquivos pelas autoras Almeida e Correia, o formato “json” é de extrema utilidade por sua estabilidade e formatação dinâmica para corpus multilíngue, como é o caso da formatação dos documentos utilizados pelo Instituto Nacional Sul-Coreano de Língua Coreana para a elaboração do *Korean-Foreign Language Parallel Corpus Project* (2023).

A quinta etapa, "nomeação dos arquivos", apesar de parecer simples, desempenha um papel importante na identificação e organização dos arquivos. Essa etapa é especialmente importante para corpora de maior volume que contém centenas ou milhares de arquivos de textos. Recomenda-se que os nomes sejam transparentes, a fim de facilitar a associação do terminólogo com o meio de divulgação, o gênero discursivo, a fonte e a data de publicação (Almeida; Correia, 2008, p. 85).

Por fim, a última etapa, "anotação dos textos", embora não seja obrigatória (demonstrada no livro como “se for o caso”), oferece benefícios significativos. A anotação, também denominada de etiquetagem, pode ser dividida em duas formas: a estrutural e a linguística. A anotação estrutural envolve a marcação de informações relevantes, tanto externas (como cabeçalhos) quanto internas (como a estrutura geral dos textos, como capítulos, parágrafos, títulos, subtítulos, notas de rodapé, tabelas e figuras). A anotação estrutural também engloba elementos como citações, abreviações, referências, datas e ênfases tipográficas (formatação textual), como negrito, itálico e sublinhado. (Almeida; Correia, 2008, p. 86).

A anotação linguística pode ser realizada em diversos níveis diferentes: morfológico, morfossintático, semântico, entre outros. Uma das características mais importantes desse tipo de anotação é a forma como ela é feita, dividida entre: automática, manual, ou semi-automaticamente. A anotação feita de forma semiautomática é amplamente considerada a mais eficiente, pois possibilita uma revisão mais rápida do que a inteiramente manual e a geração de dados mais precisos do que a inteiramente automática (Almeida e Correia, 2008, p. 87).

4. METODOLOGIA

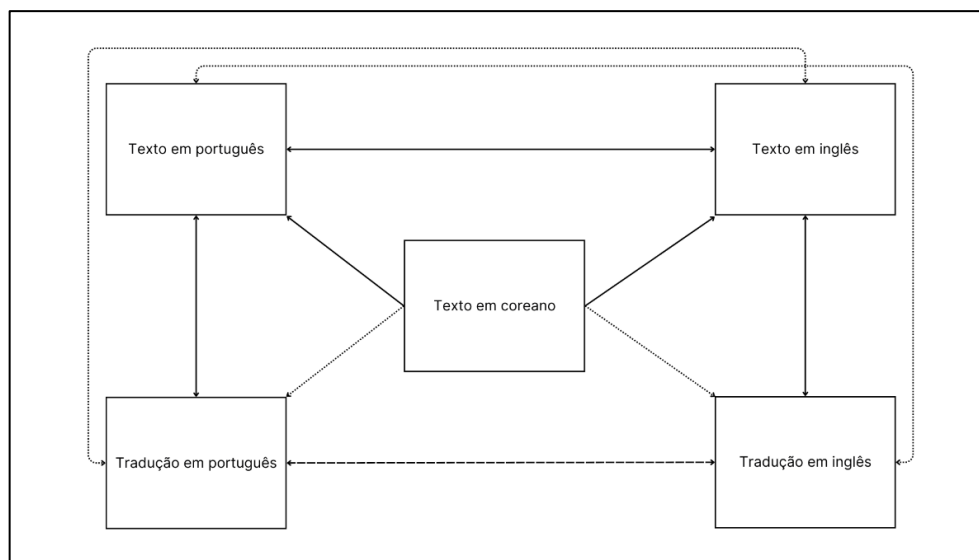
A metodologia teve como base as cinco primeiras etapas da criação de um corpus fundamentado para a análise terminológica delimitadas por Almeida e Correia (2008). A sexta etapa, das anotações, foi desconsiderada devido às limitações de tempo e fuga do escopo de criação e estabelecimento de um corpus inicial. O formato geral dos textos que serão usados para compor o nosso corpus se encaixam, em sua maioria, na perspectiva do corpus tradutório, devido à natureza dos textos originais terem a tradução realizada para outras línguas. Com o intuito de possibilitar maior completude ao corpus, também serão incluídos textos dos

documentos oficiais de jogos em coreano, inglês, e em português, a fim de tornar possível o estudo comparativo dos textos.

A elaboração do corpus é limitada essencialmente por dois fatores: tempo e disponibilidade de conteúdo. Conforme previamente comparado com o norueguês, não são abundantes os textos originais em coreano com equivalências em português e vice-versa. Tal limitação é um dos motivos para a inclusão da língua inglesa no corpus, visto que pode ser uma língua intermediária para apoiar as traduções e localizações de conteúdo em situações em que o português não é presente.

A figura 3, a seguir, toma como base a estrutura do corpus ENPC e adapta às necessidades de aplicação para o nosso corpus. Para melhor entender a figura, utilizaremos a seguinte formatação: utilização como corpus comparável entre textos originais em coreano, inglês e português e comparações com traduções de mesma língua (ver setas sólidas); utilização como corpus tradutório: Coreano-Português, Coreano-Inglês, Inglês-Português, Português-Inglês (ver setas pontilhadas); e também como corpus comparável entre traduções Português-Inglês e Inglês-Português (ver setas tracejadas).

Figura 3 – Estruturação do corpus paralelo coreano-português-inglês no domínio de *esports*



Fonte: elaboração do autor

Apesar de seguir modelo de estruturação ao ENPC, a falta de bidirecionalidade para traduções do coreano limitam os estudos tradutórios e comparativos do corpus. Além disso, há mais uma grave restrição, em que os números e tipos de textos a serem utilizados são restritos aos que foram traduzidos em ambas as direções (Johansson, 2007). A escassez de recursos

interlinguísticos entre as três línguas e restrição da área de atuação por escolha de um único tema (*esports*) serão problemas pertinentes para a criação e possível expansão futura do corpus.

Para a coleta de textos, utilizaremos diferentes fontes com diferentes contextos de uso para alcançar os diferentes usos de termos e possivelmente encontrar comparações entre a linguagem “oficial” – descrita pela documentação do jogo – e a de uso real dos jogadores.

Cada proposta temática será dividida em diferentes subcorpora, com um total de 3 subcorpora com divisões internas, com a seguinte divisão: 1) documentações de jogos; 2) transcrições da comunicação de jogadores profissionais em partidas oficiais e falas dos jogadores em transmissões ao vivo; 3) entrevistas realizadas com jogadores profissionais.

4.1. Subcorpus 1 – documentações de jogos

Para a coleta de textos do subcorpus de documentações de jogos, as bibliotecas API da Riot Games com o conteúdo base dos jogos serão utilizadas para a coleta dos dados de League of Legends e Legends of Runeterra. A biblioteca Data Dragon disponibiliza a documentação dos jogos mais populares da desenvolvedora: League of Legends, Valorant, Teamfight Tactics e Legends of Runeterra. Apesar disso, nem todos os jogos possuem informações na API sobre arquivos do jogo, o que resultou na escolha apenas dos jogos que possuíam tais informações (League of Legends e Legends of Runeterra).

Os jogos escolhidos possuem documentação em português, inglês e coreano, e não há referência de utilização de uma língua base e tradução a partir dela, então entendem-se como documentos originais com etiquetagem comparativa entre línguas.

Para a coleta da API do League of Legends e do Legends of Runeterra, foi utilizada uma chave API em dois momentos: um para extrair informações dos campeões, e outro para extrair informações dos itens. Os arquivos coletados estão em formato *.json* e apenas informações de texto relevantes foram escolhidas – excluindo fatores como quantidade de vida de personagens, ou preços e valores de itens. O arquivo em formato *.json* é frequentemente utilizado para transmitir dados por uma rede, e sua capacidade de organização por meio de *tags* em sua estrutura torna mais fácil a comparação de textos.

Os arquivos *.json* foram então divididos individualmente por campeão e item no caso do League of Legends, e por carta no caso do Legends of Runeterra. As *tags* presentes nos arquivos são idênticas em todos os idiomas, de tal forma que o mesmo texto sempre terá a mesma *tag* de identificação, o que facilita o alinhamento das entradas multilíngues do corpus.

Figura 4 – exemplo de arquivo .json em inglês do campeão Aatrox

```
{
  "id": "266",
  "name": "Aatrox",
  "title": "the Darkin Blade",
  "lore": "Once honored defenders of Shurima against the Void, Aatrox and his brethren would eventually become an even greater threat to Runeterra, and were defeated only by cunning mortal sorcery. But after centuries of imprisonment, Aatrox was the first to find freedom once more, corrupting and transforming those foolish enough to try and wield the magical weapon that contained his essence. Now, with stolen flesh, he walks Runeterra in a brutal approximation of his previous form, seeking an apocalyptic and long overdue vengeance.",
  "allytips": [
    "Use Umbral Dash while casting The Darkin Blade to increase your chances of hitting the enemy.",
    "Crowd Control abilities like Infernal Chains or your allies' immobilizing effects will help you set up The Darkin Blade.",
    "Cast World Ender when you are sure you can force a fight."
  ],
  "enemytips": [
    "Aatrox's attacks are very telegraphed, so use the time to dodge the hit zones.",
    "Aatrox's Infernal Chains are easier to exit when running towards the sides or at Aatrox.",
    "Keep your distance when Aatrox uses his Ultimate to prevent him from reviving."
  ],
  "spells": {
    "AatroxQ": {
      "name": "The Darkin Blade",
      "description": "Aatrox slams his greatsword down, dealing physical damage. He can swing three times, each with a different area of effect.",
      "tooltip": "Aatrox slams his greatsword, dealing <physicalDamage>{{ qdamage }} physical damage</physicalDamage>. If they are hit on the edge, they are briefly <status>Knocked Up</status> and they take <physicalDamage>{{ qgedgedamage }}</physicalDamage> instead. This Ability can be <recast>Recast</recast> twice, each one changing shape and dealing 25% more damage than the previous one.{{ spellmodifierdescriptionappend }}"
    }
  }
}
```

Fonte: elaboração do autor

4.2. Subcorpus 2 – transcrições da comunicação de jogadores profissionais em partidas oficiais e das falas dos jogadores em transmissões ao vivo

Para o subcorpus de transcrições da comunicação de jogadores profissionais em partidas oficiais e de falas dos jogadores em transmissões ao vivo são utilizados os canais do YouTube de três das maiores organizações da liga coreana: T1, Dplus KIA e Gen.G Esports.

Todas as três organizações participam ativamente dos campeonatos mundiais anuais e coletivamente contam com milhões de seguidores em suas redes sociais. Em especial, a T1, time de Lee Sang-hyeok, “Faker”, é a única tricampeã mundial existente até o momento e é considerada uma das organizações de maior influência dentro dos *esports*. Todos os três times publicam vídeos em suas plataformas sociais com frequência e contam com transcrições do coreano falado no vídeo e legendas em inglês. Serão extraídas as transcrições em coreano com o alinhamento da legenda equivalente em inglês, e em seguida serão realizadas as traduções para o português por meio do tradutor automático do Google Cloud Translate, com revisão manual para todas as entradas.

A maior dificuldade encontrada na obtenção dos dados foi a extração dos textos em coreano e inglês, uma vez que os vídeos do canal do YouTube da T1 possuíam uma legenda implementada diretamente no vídeo – comumente denominada de *hardcoded subtitles*. Tais legendas não estão em formatação de texto e a sua extração é realizada por meio do reconhecimento de imagens e conversão para texto com softwares de *Optical Character Recognition* (OCR). O problema se repete parcialmente para os canais do YouTube da Dplus KIA e Gen.G, pois esses possuíam a transcrição em coreano em formato de imagem na tela

(*hardcoded*), mas com legendas em inglês disponíveis pelo YouTube, o que possibilita a extração direta em texto do inglês, mas dificulta alinhamento por entrada.

Figura 5 – captura de tela de vídeo da T1



Fonte: T1 (2022)

Acima se encontra um exemplo de legenda presente em vídeo da T1 que traz a comunicação em tempo real realizada pelos jogadores durante partida do Campeonato Mundial de 2022. É possível perceber que o texto em coreano e inglês, apesar de legível de forma simples por humanos, conta com várias outras informações de texto e imagem em região adjacente. O exemplo retirado do vídeo mostra de forma clara a motivação por trás do corpus de contextualizar o uso real da linguagem do jogo, uma vez que em apenas uma frase, o jogador utilizou três unidades lexicais referentes ao jogo: 아펠 (Aphel) 노(no) 돌풍(Galeforce). A primeira palavra, Aphel, é a abreviação do nome de um dos personagens do jogo: *Aphelios* (nome igual em português e inglês). A segunda é considerada um empréstimo linguístico (Bastarrica, 2009) do “não”: “no”, do inglês. A terceira é referente ao nome de um dos itens do jogo; em inglês, “*Galeforce*”; em português, “Força do Vendaval”. Fora do contexto do jogo, a palavra isolada 돌풍 [dol-pung] possui o significado de vendaval ou redemoinho.

Para resolver o problema de extração e alinhamento do texto, seria necessário um software de OCR com capacidade de reconhecimento de caracteres coreanos (hangul) e do alfabeto latino ao mesmo tempo, e as limitações encontradas foram a baixa disponibilidade do serviço de forma gratuita ou de baixo preço, e qualidade do processamento da imagem para garantir a transcrição correta das entradas.

Inicialmente, utilizou-se o mecanismo *Tesseract*, mecanismo cujo desenvolvimento é realizado pela Google desde 2006. Apesar de obter resultados satisfatórios para o inglês, o programa mostrou-se ineficiente ou de complexidade demasiadamente elevada para o funcionamento do programa com caracteres do coreano.

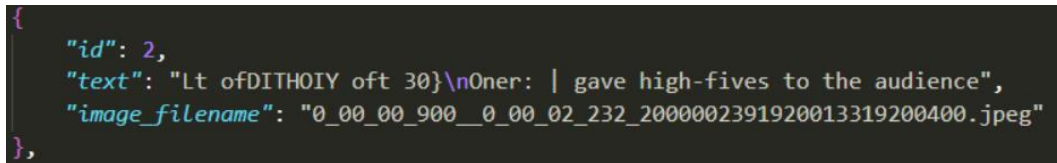
Figura 6 – captura de tela de legenda de vídeo da T1



Fonte: T1 (2022)

Após identificação por OCR com o mecanismo do *Tesseract*, o resultado foi quase inteiramente satisfatório para o português, mas o reconhecimento de caracteres do coreano foi inexistente.

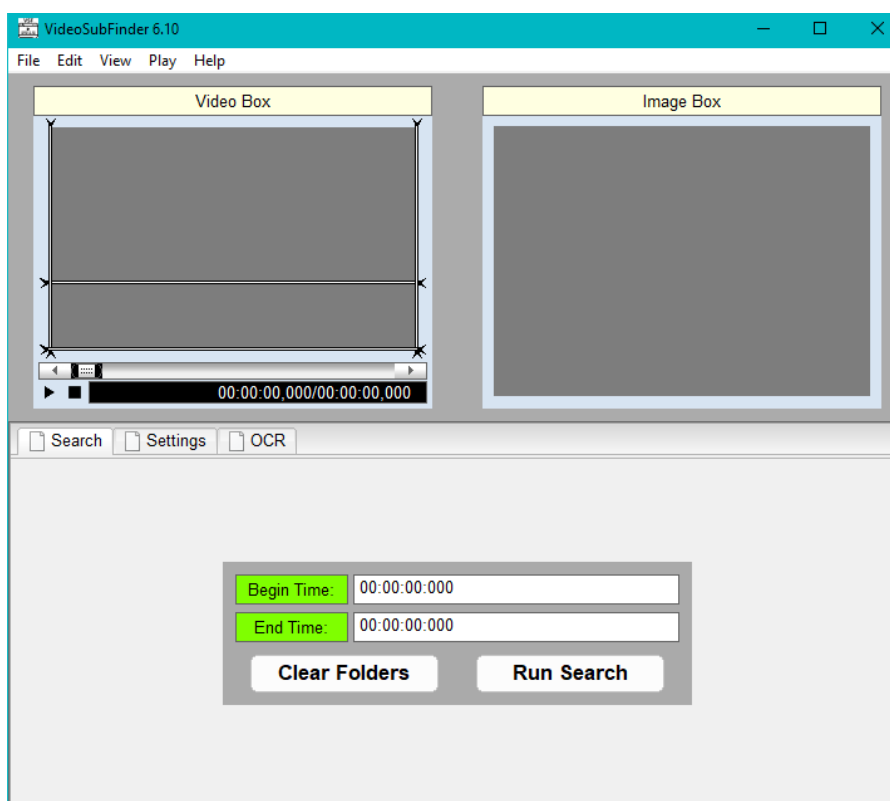
Figura 7 – transcrição realizada pelo *Tesseract*



Fonte: elaborado pelo autor

Apesar do *Tesseract* ser um mecanismo altamente customizável e com parâmetros mutáveis para todos os tipos de reconhecimento de texto, viu-se irrealista a aprendizagem específica do mecanismo para essa aplicação devido à limitação de tempo e alta complexidade de processamento.

Em seguida, a fim de gerar imagens mais limpas para um melhor processamento de mecanismos OCR, o software *VideoSubFinder*, desenvolvido por Simeon Kosnitsky em 2007 foi escolhido para realizar a identificação da legenda no vídeo e separação por imagem de cada entrada. A versão utilizada foi a 6.10, com data de atualização em 4 de abril de 2023, e o processamento de imagem realizado para cada vídeo seguiu com parâmetros específicos para garantir maior resultado no processamento das imagens.

Figura 8 – página inicial do software *VideoSubFinder*

Fonte: VideoSubFinder (2007)

O software consiste em receber um arquivo de vídeo com legenda e realizar processamentos de imagem sob o vídeo para gerar arquivos individuais do pós-processamento, podendo estes serem utilizados juntos a softwares de OCR para sua transcrição em texto, formatação temporal de entrada e saída de legendas, ou a combinação das duas finalidades para gerar um novo arquivo de legenda com marcações de tempo e entradas preenchidas pelo reconhecimento de texto por imagem.

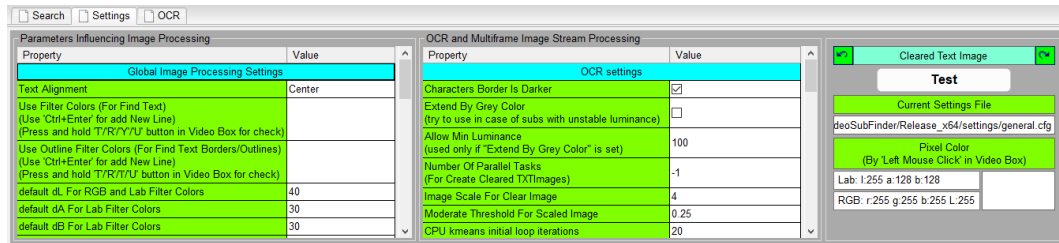
Devido a formatação das legendas nos vídeos da Dplus KIA e Gen.G, a segunda funcionalidade do programa, apesar de não ser direcionada exatamente para o que queremos, mostrou-se útil para facilitar o alinhamento com o arquivo das legendas extraído do YouTube com marcações de tempo.

Apesar dos parâmetros serem diferentes para cada vídeo analisado, a ordem de comandos manteve-se a mesma, somente com divergências em relação a exportação final (caso necessárias as marcações de tempo) e parâmetros de processamento de imagem, variável por cor e estilo da fonte de cada vídeo.

Primeiro, as barras pretas presentes na *Video Box* delimitam o local de aparição das legendas e limitam o reconhecimento de texto do software somente à área aplicada. No geral, o enquadramento das legendas manteve-se constante entre os vídeos obtidos, com exceção de

alguns vídeos da T1 que possuíam legendas em posições diferentes durante o decorrer do vídeo. Nesse caso, uma janela maior e processamento mais rígido foi escolhido.

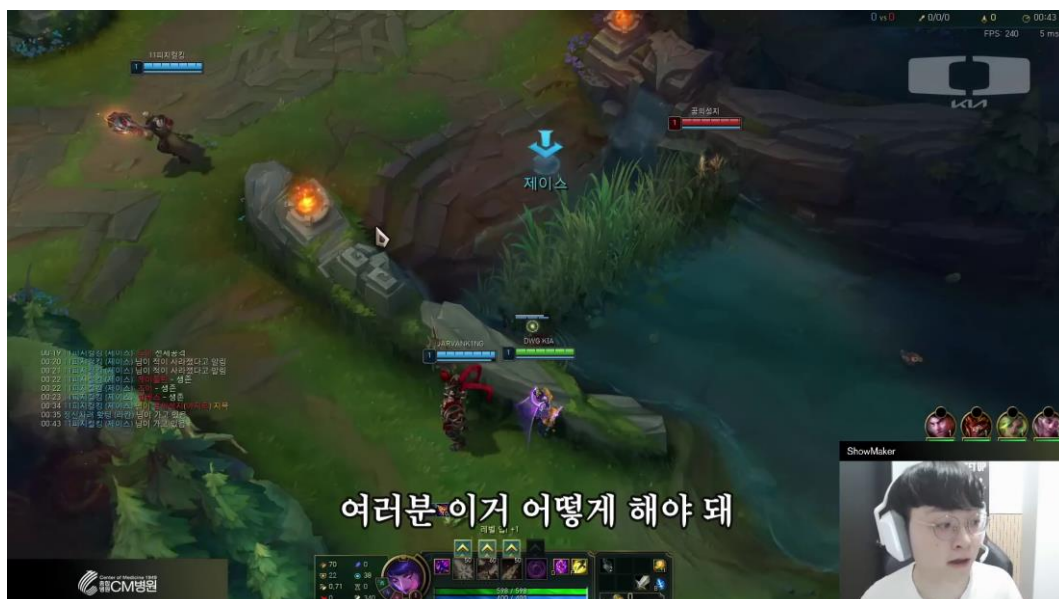
Figura 9 – janela de configurações do *VideoSubFinder*



Fonte: VideoSubFinder (2007)

A janela de configurações do software, apesar de inicialmente parecer complexa, requer poucas mudanças para casos de vídeos como os nossos – de alta qualidade e com legendas de alta luminosidade e contraste. Foram utilizados filtros de cores para diferenciação da legenda referente a caracteres ao fundo; *threshold*, referente a limite, ou tolerância, de possíveis caracteres identificados como legenda com valores entre 0.3 e 0.6 dentre uma escala de 0-1; marcação da opção de *Characters Border is Darker* no caso de legendas com borda preta e preenchimento branco. Todas as outras configurações foram mantidas conforme configuração padrão do aplicativo. Abaixo, a entrada da legenda “여러분 이거 어떻게 해야 돼 [yeoreobun igeo eotteoke haeya dwae]”, em português, “pessoal, o que devemos fazer?”, é utilizada como exemplo de extração de uma entrada realizada pelo aplicativo.

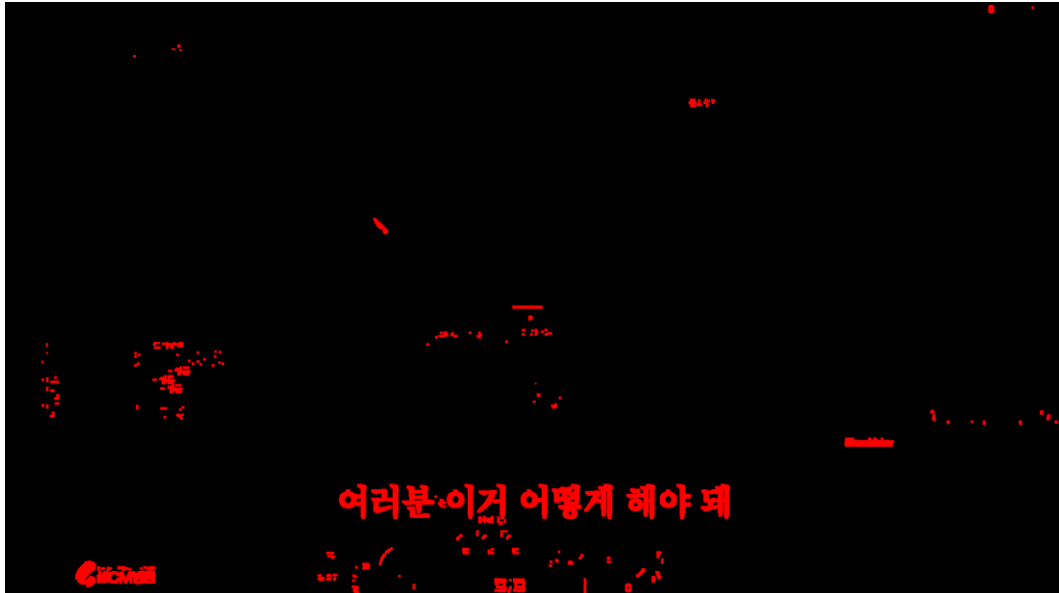
Figura 10 – imagem base para processamento no *VideoSubFinder*



Fonte: Dplus KIA (2023)

O software primeiro analisa frames do vídeo em que sua posição é constante e com duração maior de 0.5s dentro da área delimitada, em busca de possíveis legendas.

Figura 11 – imagem com processamento de filtro de cores inicial



Fonte: Dplus KIA (2023)

Após filtragem realizada por cores similares à(s) declarada(s) como preenchimento da legenda, é realizada uma segunda filtragem que destaca somente cores compatíveis.

Figura 12 – imagem com processamento de filtro de cores com bordas

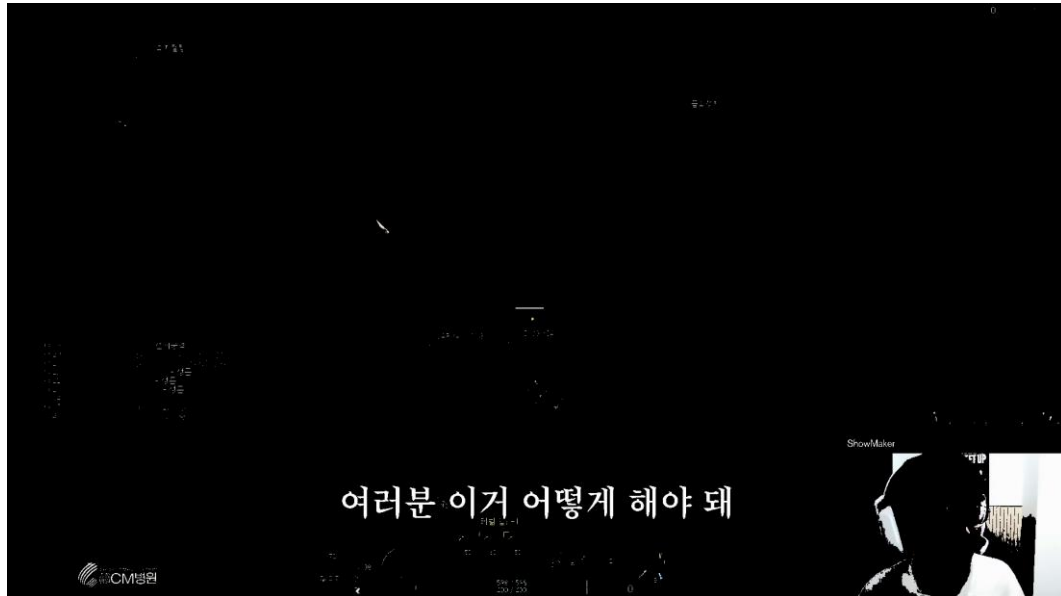


Fonte: Dplus KIA (2023)

Em seguida, no caso de legendas com borda escura, que é o caso do vídeo, o programa busca por cores que foram encontradas na etapa anterior que estão cercadas em sua maioria por

cores declaradas como borda da legenda. Dessa forma, possíveis bordas são destacadas de verde, e possíveis legendas são destacadas de vermelho.

Figura 13 – imagem com processamento de cores final



Fonte: Dplus KIA (2023)

Por fim, o software utiliza das informações anteriores para gerar uma imagem final que será posteriormente utilizada para detecção de caracteres dentro do próprio software *VideoSubFinder*.

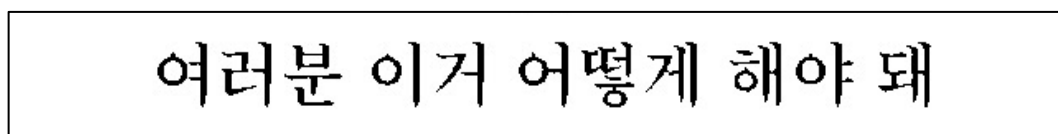
Figura 14 – pastas de arquivos de imagem geradas pelo *VideoSubFinder*

bitmaps	28/3/23 6:59 PM	File folder
DebugImages	4/10/23 3:53 PM	File folder
Docs	14/4/23 12:02 AM	File folder
ILAIImages	21/10/23 2:16 AM	File folder
ImagesJoined	21/10/23 2:36 AM	File folder
ISAIImages	21/10/23 2:16 AM	File folder
RGBImages	21/10/23 2:16 AM	File folder
settings	25/4/23 1:23 PM	File folder
TestImages	4/10/23 3:53 PM	File folder
TXImages	21/10/23 2:34 AM	File folder
TXResults	21/10/23 2:36 AM	File folder

Fonte: elaboração do autor

Por ser um programa multifacetado, são disponibilizadas ao final do processo todas as imagens geradas em todas as etapas do processamento, mas as que nos interessam para a confecção do corpus são aquelas presentes na pasta “TXImages”, que convertem a legenda, encontrada em imagem com contraste claro:

Figura 15 – exemplo imagem final da pasta TXTImages



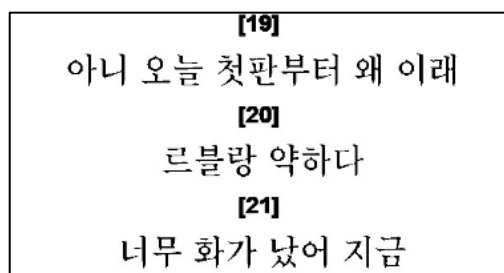
Fonte: elaboração do autor

Por fim, chega o momento de transformar a imagem com texto mais facilmente reconhecível em arquivo de texto para confecção do corpus. Em primeira instância, utilizou-se o serviço API do Google Cloud Vision, devido a sua alta precisão e rapidez de processamento, porém o projeto se mostrou financeiramente inviável ao reconhecer que centenas de legendas seriam processadas por vídeo e o serviço da Google possui uma cobrança de aproximadamente U\$1,50 por 1000 entradas. Apesar de acessível em presente momento, pensando na escalabilidade do corpus e de projetos similares, uma outra solução foi necessária.

O mesmo problema de restrições financeiras foi encontrado em plataformas API de excelente rendimento, como o Baidu AI Cloud – OCR, e o Clova OCR da NAVER, ambos com *datasets* especializados para caracteres coreanos. Além disso, ambos possuíam documentação escassa em inglês, e restringiam acesso para usuários fora do país de origem, respectivamente China e Coreia do Sul.

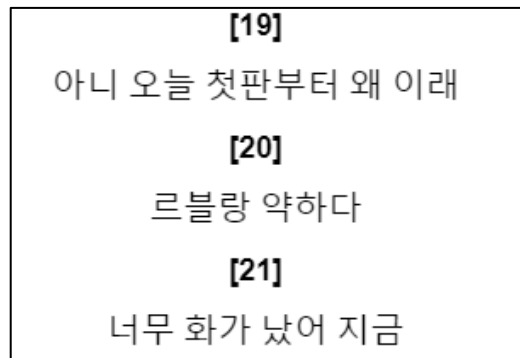
Após contemplar realizar o alinhamento de entradas de forma manual, uma das outras funções do *VideoSubFinder* mostrou-se extremamente útil para tornar possível o processo: a junção de imagens em arquivo único.

Figura 16 – recorte de arquivo único da pasta ImagesJoined



Fonte: elaboração do autor

Após realizar a junção em arquivo único, a imagem pode ser enviada ao Google Drive e aberta com o software *Google Docs*, que possui OCR nativo de todas as línguas disponíveis no ecossistema Google, e possui alto nível de precisão para imagens com informações simples, como presente em nossos arquivos.

Figura 17 – arquivo após OCR realizado pelo Google Docs

Fonte: elaboração do autor

Com o processamento OCR da imagem, basta realizar o alinhamento das entradas em inglês para os vídeos que não possuem a legenda explícita na tela – caso da Dplus KIA e Gen.G – e realizar a conferência manual das entradas ao final do processo.

Ao final, foram 8 vídeos divididos entre os três canais, com 4 deles direcionados à transcrição dos áudios durante partidas profissionais, e 4 direcionados a transmissões ao vivo e produções de conteúdo.

Tabela 2 – coleção de vídeos subcorpus 2

Nome original do vídeo	Nome em inglês do vídeo	Time
『미드』 바이가 너무 강함	Worlds Isn't Ready for Mid Vi	Dplus KIA
『월석 재생기 이즈리얼』의 비밀	The Secret of Moonstone Renewer Ezreal	Dplus KIA
GEN T1 KT 그리고 DK 레츠 고	GEN T1 KT and DK Let's GO	Dplus KIA
코그모폭탄	Icathian Surprise	Dplus KIA
새로운 왕조의 탄생 2023 LCK 서머 스플릿 결승전 Voice Comms [GEN vs T1]	2023 LCK Summer Split Grand Finals Voice Comms [GEN vs T1]	Gen.G
Embers ignited [T1scord Ep.15]	Embers ignited [T1scord Ep.15]	T1
vs JDG, 그리고 굿바이 2022 T1 Worlds Comms [T1 Hidden Track S2 EP.21]	vs JDG, 그리고 굿바이 2022 T1 Worlds Comms [T1 Hidden Track S2 EP.21]	T1
칼과 창 제우스 [T1 Stream Highlight]	Stepping Up in 4th, OP Power: Zeus' Poppy! [T1 Stream Highlight]	T1

Fonte: elaboração do autor

4.3. Subcorpus 3 – entrevistas realizadas com jogadores profissionais

Para a realização do subcorpus com entrevistas realizadas pelos jogadores profissionais, foram escolhidos conteúdos trazidos pelo canal midiático *KORIZON Esports*. O canal é conhecido por trazer entrevistas e conteúdo da Coreia do Sul para o público internacional por meio da tradução e legendagem pelo inglês. Co-fundada por Ashley Kang, a página conta com artigos e publicações constantes desde 2019 da liga oficial de League of Legends coreana – a LCK – e conteúdo dos campeonatos mundiais.

Outra razão para a escolha do canal é a qualidade e facilidade de acesso às legendas, uma vez que ambas as transcrições do coreano e legendas equivalentes são disponibilizadas diretamente pela plataforma do YouTube. Dessa forma, o alinhamento de entradas é realizado de forma mais simples e há uma mistura do uso de linguagem de especialidade utilizada por jogadores com o léxico geral utilizado nas entrevistas a fim de atingir um público maior.

Os vídeos selecionados para a inclusão inicial no subcorpus são os referentes ao Campeonato Mundial de 2023, em entrevistas realizadas com jogadores nativos da Coreia do Sul. Ao total foram selecionadas 8 entrevistas realizadas por repórteres e jogadores coreanos na língua coreana. Uma tradução manual foi realizada para o português com base nos textos da legenda do inglês e do coreano. Os arquivos foram salvos como .txt e divididos por time, jogador, data de publicação e idioma do arquivo.

Tabela 3 – coleção de entrevistas do subcorpus 3

Time	Jogador	Data
DK	Canyon	21/10/2023
DK	Canyon	29/10/2023
DK	Deft	27/10/2023
DK	Showmaker	23/10/2023
GEN	Doran	21/10/2023
KT	Cuzz	29/10/2023
KT	Lehends	27/10/2023
T1	Gumayusi	28/10/2023

Fonte: elaboração do autor

Apesar do canal da *KORIZON Esports* ser um dos poucos meios em que há uma tradução para o coreano em entrevistas realizadas com jogadores falantes do inglês, a falta da legenda em inglês e tempo requerido para realizar manualmente a transcrição dos vídeos foram

fatores determinantes para não incluir vídeos em que a língua de origem seria o inglês com a tradução para o coreano. Apesar disso, é uma fonte interessante a ser utilizada futuramente em outra oportunidade para adicionar, dessa forma, bidirecionalidade à pesquisa.

Sendo assim, os mais de mil vídeos do canal podem servir futuramente como um possível corpus bidirecional da língua coreana e da língua inglesa. A KORIZON *Esports* é uma das pioneiras na internacionalização do conteúdo coreano para o inglês e vice-versa (do internacional para o coreano), e rapidamente consolidou-se como um dos agentes midiáticos mais respeitados pela comunidade de jogos. Tal sucesso serve como indicativo da possibilidade de expansão de conteúdos na área, uma vez que apesar de ser um conteúdo que é de interesse público, muitas vezes o acesso era barrado pela barreira linguística.

4.4. Compilação e análise do corpus

Para a análise do corpus, foi utilizado o software AntConc, desenvolvido pelo Dr. Laurence Anthony em sua versão 4.2.4, publicada em 2023. O maior motivador para a escolha do AntConc é a sua gratuidade e documentação de fácil acesso para uma melhor experiência do usuário. Além disso, para ajudar na visualização das entradas alinhadas, o software Excel, da Microsoft, em versão 2310, publicada no ano de 2023, foi utilizado. As planilhas do Excel nesse caso assumiram uma função de ajuda para visualização de dados entre os diferentes idiomas, e a ferramenta da Microsoft não foi utilizada em nenhum momento para a contagem de tokens, types ou ranqueamento de palavras devido a maior adaptação do AntConc para tal tarefa.

Em relação ao ranqueamento de palavras e identificação de termos, a biblioteca NLTK – *Natural Language Toolkit* – foi utilizada em sua versão 3.8.1, publicada em 2023, para a filtragem de *stopwords* – palavras de uso comum classificadas como não importantes/sem conteúdo único, como artigos, conectores e similares – do português e do inglês. O processamento dos arquivos foi feito por meio de um script no Python e com a utilização das bibliotecas “tokenize.word_tokenize” e “corpus.stopwords” do NLTK para a remoção de *stopwords* ao criar listas de palavras mais frequentes, buscando facilitar o encontro de candidatos a termo. Pela falta de suporte à língua coreana dentro da biblioteca NLTK, foi escolhido repositório de *stopwords* da língua coreana criado por Gene Diaz em 2016 e disponível por meio da plataforma GitHub¹.

¹ Disponível em: <https://github.com/stopwords-iso/stopwords-ko/commits?author=genediazjr> (acesso em 4 de dez. 2023)

Figura 18 – função no Python para remoção de *stopwords* no ranqueamento de palavras

```
def count_word_frequency(dataframe, column_name):
    tokens = []
    for row in dataframe[column_name].dropna():
        tokens.extend(word_tokenize(row.lower())) # Tokenize and convert to lowercase

    # This code is to remove stopwords/punctuations in English/Portuguese data using NLTK library
    # If you use this code, comment out the code for Korean data, and vice versa
    stop_words = set(stopwords.words('english'))
    tokens = [token for token in tokens if token.isalnum() and token not in stop_words]

    # This code is to remove stopwords/punctuations in Korean data as the language is not covered in NLTK library
    tokens = [token for token in tokens if token not in stop_words.ko.json]

    most_frequent_words = Counter(tokens).most_common(50) # Count the 50 most frequent words
    return most_frequent_words
```

Fonte: elaboração do autor

A figura mostra a função utilizada no Python para a remoção de *stopwords* no português e no inglês por meio da biblioteca NLTK, e o mesmo processo realizado separadamente para a língua coreana por meio do arquivo disponibilizado no repositório de Gene Diaz.

5. RESULTADOS

O resultado da coleção dos três subcorpora foi um corpus de **297684 tokens** e **36128 types**.

Tabela 4 – número de tokens e types do corpus

<i>Língua</i>	Token	Type
<i>Coreano</i>	80379	19219
<i>Inglês</i>	105696	8092
<i>Português</i>	111609	9733

Fonte: elaboração do autor

É possível perceber inicialmente uma grande diferença na quantidade de tokens e types entre a língua portuguesa e a língua coreana. Isso é um padrão recorrente em todos os subcorpora e uma possível explicação é a característica aglutinante da língua coreana, que por meio de partículas e indicadores pode restringir ideias diversas em uma única unidade lexical. Além disso, a língua coreana pode ser classificada como uma língua de *pro-drop*, referente a *pronoun dropping*, que é uma língua em que certas classes de pronomes podem ser omitidas quando podem ser inferíveis pragmática ou gramaticalmente (Welo, 2013). Ambas as características citadas podem ser vistas em uma das entradas do subcorpus 3 na entrevista do Canyon da data 21/10/2023 (arquivo “DK_Canyon_20231021.txt”):

- Texto original em coreano: 팬분들한테 메세지요? [*paenbundeulhante mesejiyo?*]
 - Texto em inglês: A message to fans?!
 - Texto em português: Uma mensagem para os fãs?!

A mesma entrada conta então com 2 tokens no coreano, e 5 no português. Primeiramente, no português, a direcionalidade em “para fãs” é substituída pela partícula “한테” [*hante*], que traz a ideia de “para algo/alguém” e a pluralidade, em “os fãs” é realizado pelo caractere “들” [*deul*], que traz a ideia de plural. Além disso, o “요” [*yo*] no final indica nível de polidez na fala, que apesar de não ter tradução exata para o português e o inglês, gera diferentes *tokens* a depender da terminação formal ou informal utilizada. Devido a esta e outras situações em que o sistema de partículas e caracteres aglutinantes do coreano diminui a quantidade de itens lexicais, o corpus da língua coreana é significativamente menor em quantidade de *tokens*.

Após a compilação individual dos três subcorpora, foi criada uma tabela na ferramenta Excel com abas individuais para cada corpus e aba geral com a coleção dos corpora, com o intuito de facilitar a identificação comparativa de entradas. Dessa forma, foi possível identificar variações entre as línguas em uma perspectiva *intracorpora*, e comparação do uso de terminologias do jogo, como itens e campeões, em uma perspectiva *intercorporata*. Em presente estudo, serão demonstrados exemplos encontrados no corpus para demonstrar padrões encontrados em ambas as perspectivas.

Em tabelas com caracteres coreanos serão incluídas romanizações (entre colchetes) em suas devidas colunas para facilitar a leitura dos que não possuem familiaridade com o hangul.

5.1. Análise intracorpora

O subcorpus 1 conta com 250974 tokens e 31599 types distribuídos entre 694 arquivos, referentes aos 165 campeões e 158 itens existentes no League of Legends e 371 cartas existentes no Legends of Runeterra.

Tabela 5 – número de tokens e types subcorpus 1

<i>Língua</i>	Token	Type
<i>Coreano</i>	69284	16214
<i>Inglês</i>	87190	9087
<i>Português</i>	94500	10176

Fonte: elaboração do autor

Tabela 6 – 25 types mais frequentes subcorpus 1

Inglês		Coreano		Português	
Type	Frequência	Type	Frequência	Type	Frequência
damage	1740	피해를 [pihaereul]	1579	dano	1885
seconds	833	수 [su]	767	inimigos	970
enemies	816	마법 [mabeop]	736	s	775
magic	734	입힙니다 [iphipnida]	680	causando	701
enemy	674	초 [cho]	654	mágico	641
dealing	631	기본 [gibon]	467	inimigo	525
attack	468	있습니다 [itseupnida]	460	velocidade	515
speed	437	적에게 [jeokege]	429	vida	493
health	409	시 [si]	404	ataque	468
target	397	적을 [jeokeul]	403	causa	454
hit	364	이동 [idong]	390	habilidade	426
ability	355	입히고 [iphigo]	379	pode	424
champion	339	적 [jeok]	377	alvo	421
deals	335	공격 [gonggyeok]	351	tempo	339
physical	318	있는 [itneun]	329	físico	325
move	299	스킬을 [seukileul]	327	próximos	318
nearby	276	물리 [mulri]	318	adicional	318
attacks	261	속도가 [sokdoga]	300	campeão	317
gains	258	증가합니다 [jeungghapnida]	281	movimento	286
champions	244	추가 [chuga]	260	lentidão	284
first	222	재사용 [jaesayong]	235	campeões	263
bonus	191	최대 [choedae]	221	ataques	249
time	187	후 [hu]	209	recebe	247
shield	187	모든 [modeun]	209	cada	214
second	186	체력을 [cheryeokeul]	177	habilidades	208

Fonte: elaboração do autor

O segundo type mais frequente na língua portuguesa, “s”, se dá pelo fato de que uma grande parte do subcorpus 1 é referente às habilidades dos campeões de League of Legends, e parte das explicações das habilidades incluem a quantidade de segundos de duração de uma habilidade. A escolha estilística na documentação portuguesa foi a de abreviar segundos para

“s”, enquanto a documentação da língua coreana e da língua inglesa mantiveram as palavras completas: 초 [cho] (quinta posição) e *seconds* (segunda posição).

Assim como os outros subcorpora, o número de tokens da língua coreana é inferior aos da língua portuguesa, porém o sistema de partículas e aglutinações da língua coreana é mais evidente no subcorpus 1, uma vez que a linguagem utilizada é de caráter mais formal e há a necessidade de explicação detalhada de habilidades, histórias e características de elementos do jogo.

É interessante também realizar uma pequena divisão dentro do próprio subcorpus 1, entre informações do jogo Legends of Runeterra, e informações dos campeões e itens do League of Legends. Por se tratar de um jogo de cartas que é jogado em turnos, o Legends of Runeterra depende de uma maior descrição de ordem dos acontecimentos e capacidades de cada carta. Enquanto isso, o escopo terminológico aparenta ser muito maior quando se trata de League of Legends, um jogo com interações entre jogadores, itens, terreno e diversos outros fatores em tempo real.

Uma das palavras mais frequentes nesse subcorpus é a referente a dano, no sentido de dano a ser causado ou sofrido por personagens. Esse dano pode ser acompanhado de vários adjetivos para descrever o seu tipo e origem (como por exemplo: mágico, físico, real). Uma das características de maior relevância dentro dessa temática de dano é a escolha do uso da palavra 피해 [pihae] como equivalente de “dano” no português (“*damage*” no inglês).

A escolha do substantivo 피해 [pihae] é interessante por diferentes fatores. Em primeiro lugar, pelo fato de muitos jogos possuírem exclusivamente a língua inglesa em sua documentação, existe hoje no léxico coreano a palavra 데미지 [demiji] uma espécie de estrangeirismo existente na forma da adaptação da palavra *damage* do inglês para o alfabeto coreano – fenômeno esse que chamaremos de “hangulização”. Além disso, há também um termo designado especialmente para o ato de *causar dano*, possivelmente proveniente do verbo do inglês, *deal* (damage): 딜 [dil]. Ambos os termos vindos do inglês foram utilizados no subcorpus 2, e uma das cinco posições do League of Legends também utiliza a sua forma proveniente do inglês, no termo 윈딜 [wondil], referente à atirador, ou àquele que causa danos à distância.

Apesar de não saber exatamente o motivo pelo qual a escolha de utilizar uma palavra exclusiva do léxico coreano, é possível imaginar algumas possibilidades: devido a necessidade

de maior clareza e acessibilidade no ambiente de documentação que servirá como base do jogo; devido ao desejo de preservar o léxico coreano sem empréstimos ou estrangeirismos do inglês. Outra observação importante sobre o substantivo é a existência de uma palavra de grafia similar existente em forma verbal, 피하다 [*pihada*], com o significado de desviar; esquivar. O verbo também se torna uma homonímia perfeita em conjugação no presente declarativo informal: 피해 [*pihae*].

Apesar da forma verbal ter o uso mais frequente no sentido de esquivar ou desviar (conforme presente no subcorpus 2), a escolha lexical para tal conceito foi a de 호피 [*hoepi*] em 3 das 9 ocorrências do lema “esquivar”, do português. 호피 [*hoepi*] possui a tradução no Naver Korean-English Dictionary (Naver, 2023) como *evasion* ou *avoidance*, o que leva a significados ligeiramente diferentes do vocabulário de “esquiva” escolhido pela versão português, ou “*dodge*”, do inglês. As outras 6 aparições contam com diferentes conjugações do verbo 피하다 [*pihada*].

Para uma análise mais profunda em relação a esse termo específico, seria necessário a comparação com documentações de diferentes jogos para entender melhor as motivações estilísticas por trás de cada item lexical, uma vez que a mecânica de esquiva não é presente no Legends of Runeterra, o que impossibilita a comparação com itens dentro do corpus.

O subcorpus 2 conta com 26.028 tokens e 5.158 types distribuídos entre 24 arquivos, referentes aos 8 vídeos divididos entre os três canais citados anteriormente.

Tabela 7 – número de tokens e types subcorpus 2

<i>Língua</i>	Token	Type
<i>Coreano</i>	6368	2349
<i>Inglês</i>	10293	1647
<i>Português</i>	9367	1964

Fonte: elaboração do autor

Tabela 8 – 25 types mais frequentes subcorpus 2

Inglês		Coreano		Português	
Type	Frequência	Type	Frequência	Type	Frequência
flash	122	이거 [igeo]	107	flash	121
let	104	돼 [dwae]	99	vamos	113
ult	84	안 [an]	59	vou	92
good	76	노플 [nopeul]	57	ult	74
go	68	한 [han]	49	podemos	62
look	67	궁 [gung]	46	ei	59
hey	60	수 [su]	44	aqui	56
faker	58	천천히 [cheoncheonhi]	44	bom	55
oner	56	있어 [iteo]	43	pode	52
back	53	나이스 [naiseu]	42	posso	50
got	52	아펠 [apel]	38	olha	48
going	49	거 [geo]	36	parece	47
nice	46	미드 [mideu]	34	fazer	43
zeus	46	해 [hae]	34	jogo	43
keria	46	볼게 [bolge]	34	faker	43
fight	44	번 [beon]	32	ah	41
right	43	진짜 [jinjja]	31	aphelios	40
game	42	봐줘 [bwajwo]	29	bem	39
t1	40	잘 [jal]	28	zeus	39
mid	38	바론 [baron]	28	t1	38
get	37	그냥 [geunyang]	28	oner	37
one	37	싸움 [ssaum]	26	agora	33
take	36	중 [jung]	26	ok	31
need	35	아트 [ateu]	25	legal	31
keep	35	너무 [neomu]	24	mid	30

Fonte: elaboração do autor

O subcorpus 2 é o mais importante de todo o corpus, visto que é por meio dele que é possível realizar a desejada comparação entre diferenças da terminologia utilizada em meios de comunicação de caráter mais oficial e do uso real entre jogadores. Apesar disso, pela necessidade de tradução para o português e dificuldade na extração de legendas, é desejável uma quantidade e variedade ainda maior de informações para abranger jogadores da comunidade que não sejam profissionais, ou conteúdos mais variados.

Além disso, apesar de haver a divisão entre três times relevantes do cenário de *esports* sul-coreano, a inclusão de ainda mais times seria interessante, mas muitos sequer possuem a tradução para o inglês, que possibilitaria uma adaptação mais fácil ao português.

Tal limitação pode ser claramente vista na análise dos tokens, em que a palavra 우제 [uje], nome próprio masculino possui apenas uma equivalência: “Zeus”. Isso ocorre não pelo fato de o nome em coreano significar Zeus, mas pelo jogador participante da equipe da T1 ser a única pessoa do corpus com esse nome, e o seu *nickname* no jogo ser Zeus. Esses e outros exemplos podem ser vistos através de todo o corpus, como a presença do próprio nome da equipe, T1, aparecendo como token frequente, ou a presença dos nomes de outros jogadores dentre os 25 tokens mais frequentes: Faker, Keria, Oner.

A falta de conteúdos em português com tradução para o inglês também é outro fator importante a ser considerado, uma vez que não é possível ver como seria feita a adaptação da comunicação de jogadores da comunidade brasileira para a língua coreana apenas com os dados presentes.

O subcorpus 3 conta com 20.731 tokens e 4.497 types distribuídos entre 24 arquivos, referentes à 8 entrevistas realizadas em 2023 na língua coreana no canal do YouTube da KORIZON *Esports*, conforme citado anteriormente.

Tabela 9 – número de tokens e types subcorpus 3

<i>Língua</i>	Token	Type
<i>Coreano</i>	4759	2062
<i>Inglês</i>	8220	1168
<i>Português</i>	7752	1464

Fonte: elaboração do autor

Tabela 10 – 25 types mais frequentes subcorpus 3

Inglês		Coreano		Português	
Type	Frequência	Type	Frequência	Type	Frequência
dk	59	같아요 [gatayo]	48	sobre	64
worlds	58	그런 [geureon]	47	jogo	64
game	53	많이 [mani]	43	dk	58
think	52	잘 [jal]	36	mundial	56
team	37	수 [su]	35	então	43
kt	35	오늘 [oneul]	29	kt	35
lane	35	저희가 [jeohuiga]	29	final	34
good	34	같습니다 [gatseupnida]	27	bem	33
like	33	같은 [gateun]	25	equipe	31
really	31	할 [hal]	24	contra	30
also	30	게 [ge]	23	acho	30
today	30	있는 [itneun]	22	hoje	30
right	28	더 [deo]	21	scrim	26
playing	28	하고 [hago]	20	dizer	26
would	26	대한 [daehan]	20	ter	24
well	25	대해 [daehae]	20	desempenho	24
one	25	너무 [neomu]	19	quartas	24
stage	25	좋은 [joteun]	19	penso	24
quarterfinals	24	같고 [gatgo]	18	quero	24
performance	23	q [q]	18	melhor	23
us	23	한 [han]	17	lane	22
get	23	제가 [jega]	17	bom	22
face	21	생각해서 [saenggakhaeseo]	17	porque	21
want	21	선수가 [seonsuga]	17	realmente	21
feel	21	하는 [haneun]	16	vamos	20

Fonte: elaboração do autor

As entrevistas foram especialmente úteis para visualizar a tradução de termos que vão além do decorrer das partidas, em especial a fase de escolha dos campeões que ocorre previamente às partidas de League of Legends. Há a presença de um empréstimo vindo do inglês para o português para se referir a essa fase, por meio da palavra *draft* – referente a esboço;

projeto. O ato de selecionar e banir campeões antes do início da partida ficou popularmente conhecido em português como o ato de “draftar”, com uma frequência do radical “draft” aparecendo ao total 34 vezes no subcorpus. Por meio do alinhamento de entradas, foi possível traçar uma relação tradutória do termo com o termo coreano de 밴픽 [baenpik], que também é de origem inglesa, de *ban* e *pick*, seguindo a ideia de escolhas e banimentos. Os termos de escolha e banimentos de campeões nessa fase pré-jogo também penetraram para a língua portuguesa, com o radical “pick” aparecendo 15 vezes no subcorpus.

Outro termo presente no corpus que permeia todas as línguas é a palavra *scrim*, do inglês: significando partidas-treino jogadas com outras equipes como forma de treino ou prática. O termo teve a sua adaptação tanto como substantivo como passou por um processo de “verbalização” no português, assim como outros empréstimos. O termo também passou pelo processo de “hangulização” e assumiu a forma de 스크림 [seukeurim]. A permeabilidade do termo pode ser vista em entrada a seguir, retirada do subcorpus 3:

- Texto original em coreano: 이번에 스크림도 몇 번 해보고 하면서 [ibeone seukeurimdo myeot beon haebogo hamyeonseo]
 - Texto em inglês: But after scrimming against them a few times,
 - Texto em português: Mas depois de scrimmar com eles algumas vezes,

Uma tabela pode ser montada com algumas das equivalências terminológicas encontradas entre as línguas referentes aos elementos discutidos em entrevista que vão além do vocabulário utilizado no decorrer das partidas:

Tabela 11 – equivalência de termos do subcorpus 3

Inglês	Coreano	Português
Champion pool	챔피언 풀 [chaempieon pok]	Champion pool
Scrim	스크림 [seukeurim]	Scrim
Draft	밴픽 [baenpik]	Draft / Picks e bans / Fase de escolhas
Composition / Comp	조합 [johap]	Composição / Comp
Solo queue	솔로 랭크 / 솔랭 [solro raengkeu / solraeng]	Solo queue

Fonte: elaboração do autor

5.2. Análise intercorpora

Fica evidente a grande adaptação da linguagem utilizada por jogadores em contextos de uso real ao se comparar a terminologia utilizada para se referir a campeões, itens e feitiços do jogo. Um primeiro fenômeno evidente em todas as línguas, mas mais ainda no coreano, é a abreviação de palavras.

O campeão Aphelios, do League of Legends, apesar de aparecer todas as vezes com o seu nome completo, 아펠리오스 [*apellioseu*] na documentação presente no subcorpus 1 (frequência 10) e utilizado uma vez no subcorpus 3, não é utilizado em nenhum momento com essa grafia completa no subcorpus 2. Em contexto de uso durante as partidas ou transmissões, o nome do campeão foi adaptado em coreano para 아펠 [*apel*], referindo-se somente ao início da palavra (*Aphel*), a fim de alcançar uma comunicação mais rápida e efetiva. O termo aparece ao total 40 vezes no subcorpus 2 para a língua coreana, e a grafia completa referente ao nome do campeão não é utilizada uma sequer vez.

O caso do Aphelios é um tipo de abreviação realizada exclusivamente pela língua coreana no escopo encontrado nos corpos, uma vez que o nome do campeão é utilizado em sua integridade em todos os subcorpora da língua portuguesa. Apesar disso, o fenômeno de abreviação de nomes também existe no português para outros campeões, feitiços e itens. A seguir, é possível visualizar uma tabela com exemplos dessas abreviações presentes no subcorpus 2 e sua frequência de uso comparado com a sua forma original. Em casos da ausência de abreviação no português e existência no coreano (como Aphelios), um hífen [-] será utilizado para demonstrar tal ausência. Vale lembrar que nenhuma dessas abreviações aparece nos subcorpora 1 e 3, considerados meios mais “oficiais” de comunicação.

Tabela 12 – abreviações no subcorpus 2

Nome em PT e KR	Abreviação em PT e KR	Frequência de uso abreviado KR	Frequência de uso completo KR	Frequência de uso abreviado PT	Frequência de uso completo PT
Aphelios / 아펠리오스	- / 아펠 [apel]	40	0	-	40
Sejuani /세주아니	Seju / 세주	12	1	17	2
Renekton / 레넥톤	Renek /레넥	7	1	7	2
Flash / 점멸*	- / 플	114	2	-	116
Malphite / 말파이트	- / 말프	14	2	-	16
Aatrox / 아트록스	- / 아트	25	0	-	25
Sylas / 사일러스	- / 사일	20	0	-	23
Ultimate / 궁극기	Ult / 궁	92	1	93	-

Fonte: elaboração do autor

Destacado com um asterisco, o termo *flash*, referente a um “feitiço de invocador” dentro do jogo não só é um exemplo de abreviação, porém de um outro fenômeno de alta ocorrência para ambos o coreano e o português, que é o estrangeirismo ou empréstimo do inglês para a formação de termos nas línguas receptoras. Conforme discutido anteriormente, devido a adaptação ao léxico e eventuais mudanças ortográficas, é prudente considerar essas inclusões como empréstimos.

Os feitiços de invocador, apesar de terem traduções equivalentes para cada língua (com exceção do flash para o português e o inglês, que utilizam o mesmo nome), realizam um grande empréstimo da língua inglesa, como é o caso do *flash*, que se encontra no corpus tanto sendo usado como verbo quanto substantivo:

- Texto original em coreano: 나 플 쓸게 [*na peul sseulge*]
 - Texto em inglês: I'll **Flash** in.
 - Texto em português: Vou **flashar**.
- Texto original em coreano: 플이라도 빼보자 [*peurirado ppaeboja*]
 - Texto em inglês: Let's at least get his **Flash**.
 - Texto em português: Vamos pelo menos gastar o **Flash** dele.

A maior diferença desse estrangeirismo que acontece tanto no coreano quanto no português é a inclusão de uma etapa de abreviação que ocorre no coreano, em que o termo 플 [*peul*] é uma abreviação da grafia completa de 플래시 [*peullaesi*]. Para esse feitiço em específico, é interessante notar que ele é quase somente chamado por seu nome “oficial” de 점멸 [*jeommyeol*] em arquivos e páginas oficiais do jogo – uma frequência de 7 aparições no subcorpus 1, e 2 aparições no subcorpus 2, muito inferior às 114 ocorrências da forma abreviada do inglês.

Além disso, a expressão do inglês foi trazida para o coreano também ao indicar a ausência do flash, com a maior colocação à esquerda do termo 플 [*peul*] sendo a palavra 노 [*no*], diretamente ao “no”, “não” em inglês; ou seja, 노 플 [*no peul*] alinha-se com “sem flash” no português. Tal colocação aparece 79 vezes no subcorpus 2 e reforça a forte influência do inglês no vocabulário terminológico do coreano no meio dos *esports*.

O feitiço de invocador *teleporte* passa pelo mesmo processo de empréstimo do inglês. Citado pelo seu nome oficial em coreano, 순간이동 [*sunganidong*] 27 vezes no subcorpus 1 (documentação oficial), o feitiço é denominado de 텔 [*tel*] em uso pelos jogadores, com 32 instâncias de uso no subcorpus 2. O termo é uma abreviação do inglês de *teleport*, com a sua grafia inteira no coreano sendo representada como 텔레포트 [*tellepoteu*]. O mesmo empréstimo e abreviação de forma similar também ocorre no português, com o feitiço assumindo a forma de “TP” (21 instâncias no subcorpus 2), ou uma verbalização em “tpar” (4 instâncias):

- Texto original em coreano: 아래에 잭스 텔 정면 [*araee jaekseu tel jeongmyeon*]
 - Texto em inglês: Below, Jax **TP'ing** in front.
 - Texto em português: Embaixo, Jax **tpando** na frente
- Texto original em coreano: 나 5 초 뒤 텔 돼, 5 초 뒤 [*na 5cho dwi tel dwae, 5 cho dwi*]
 - Texto em inglês: I can **Teleport** in 5 seconds, 5.
 - Texto em português: Posso dar **TP** em 5 segundos, 5.

Um exemplo de um empréstimo do inglês que ocorreu para o português, mas não para o coreano é referente à habilidade de jogo *Ultimate*. A abreviação ocorre para ambas as línguas, com o coreano frequentemente utilizando 궁 [*gung*] para se referir ao termo 궁극기 [*gunggeukgi*], e o português utilizando “ult” para se referir à habilidade “ultimate”. No subcorpus 1, o termo é utilizado em sua grafia completa 77 vezes na língua coreana e 68 vezes no português, porém a sua utilização foi feita de forma abreviada em todas as 186 aparições (divididas em 93 para cada língua) no subcorpus 2.

Foi possível também encontrar no corpus termos – trazendo a “hangulização” do coreano para a escrita romanizada – iguais nas três línguas, especialmente os que se referem a elementos internos do jogo na linguagem utilizada pela comunidade. Ou seja, por mais que às vezes haja uma tradução para a língua alvo, como *flash* para o coreano, o uso real do termo se mostrou predominantemente diferente e similar às outras línguas. Um outro exemplo prático são as *rotas*, locais do mapa de jogo do League of Legends que se dividem em: rota superior, rota do meio e rota inferior. Apesar da tradução ter sido feita do inglês para o português e o uso da palavra “rota” no subcorpus 1 aparecer em 35 ocasiões e em 2 ocasiões no subcorpus 3, o termo *lane* – *top lane, mid lane, bottom/bot lane* – é mais abertamente aceito e utilizado pela comunidade, conforme demonstrado pela falta de aparições no corpus 2, e inclusão do termo *lane* na língua portuguesa em 37 vezes, e a os termos *top, mid e bot* somando juntos 110 ocasiões divididas entre os subcorpora 2 e 3.

A tabela a seguir mostra alguns termos que, apesar de não necessariamente serem a tradução oficial adotada pelo jogo, seguem o mesmo padrão de similaridade de uso (dado por maior frequência e equivalência no corpus) entre os três idiomas:

Tabela 13 – termos similares entre as três línguas

Coreano	Inglês	Português
미드 [<i>mideu</i>]	Mid	Mid
탑 [<i>tap</i>]	Top	Top
바텀 [<i>bateom</i>]	Bot	Bot
플래시 [<i>peullaesi</i>]	Flash	Flash
다이브 [<i>daibeu</i>]	Dive	Dive
정글 [<i>jeonggeul</i>]	Jungle	Jungle
바론 [<i>baron</i>]	Baron	Baron

Fonte: elaboração do autor

6. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi criar um corpus paralelo que introduzisse a linguagem específica de jogos do coreano à comunidade brasileira em seu contexto de uso real e servisse como referência em uma área de conteúdo escasso, e dessa forma, o objetivo foi parcialmente atingido, uma vez que demonstra a diferença entre a linguagem oficial e a utilizada em contextos reais, e estabelece um pequeno corpus inicial que pode ser expandido futuramente, mas peca em variedade conteudista e bidirecionalidade linguística para melhor análise.

Apesar da criação desse corpus demonstrar a possibilidade de análise linguística entre as três línguas, ele é, em sua natureza, um corpus com muitas limitações. A maior delas, conforme vista no processo de coleção e compilação de textos, é a falta de bidirecionalidade entre as línguas, em que é extremamente limitado o conteúdo de língua portuguesa do meio que seja traduzido para a língua coreana. Além disso, a própria falta de adaptação dos conteúdos coreanos diretamente para a língua portuguesa e a necessidade da inclusão da língua inglesa como idioma intermediário demonstra a necessidade de expansão do domínio para possibilitar maior interação interlinguística.

Assim como é o caso de muitos outros corpora, o fator do tempo também é um grande limitante em relação à qualidade e quantidade dos textos, uma vez que a tradução manual para o português e coleta de legendas *hardcoded* demandam uma grande quantidade de tempo, além de possibilitar inconsistências e possivelmente criar parcialidades no corpus. O presente corpus

também é limitado do ponto de vista de variedade de recursos, de tal forma que apenas 3 times, 1 canal de notícias e 1 empresa de jogos foi contemplada para a sua criação, o que gera um corpus possivelmente enviesado.

Apesar disso, o presente trabalho se mostrou satisfatório em demonstrar que há uma grande diferença entre a linguagem considerada “oficial”, ou seja, provinda dos manuais e arquivos dos jogos e a de uso real pela comunidade de jogadores. Em ocasiões futuras, é desejável a criação de um corpus ainda mais diverso e preciso para possibilitar a comparação entre línguas dentro do domínio dos *esports*. A inclusão de traduções para o coreano de conteúdos em português seria um excelente começo para impulsionar essa cooperação linguística intercultural e possibilitar maior entendimento por parte da comunidade brasileira no universo dos *esports* coreanos e vice-versa.

Espera-se que o corpus sirva de referência para pesquisas futuras na área, mesmo que não usado em sua forma íntegra, mas que assuma a posição de agente motivador para demonstrar que há interesse no desenvolvimento de uma maior cooperação dentro do domínio dos *esports* e de que é possível tornar mais acessível um nicho terminológico que muitas vezes foge dos livros tradicionais de ensino de línguas. A sua aplicabilidade prática pode ser tanto em nível profissional, para os profissionais da área utilizarem como possível glossário ou referência ao incluir jogadores sul-coreanos em times falantes do português ou inglês, quanto para o uso de criadores de conteúdo, possibilitando uma legendagem mais rápida, prática, e até mesmo possivelmente automatizada – por meio de treinamento de tradutores automáticos com o corpus como *dataset*.

Por fim, apesar do corpus ter sido utilizado apenas para uma breve análise linguística e encontro de possíveis termos entre as línguas, a sua dinamicidade e capacidade de expansão possibilitam o seu uso como um *dataset* multilíngue para futuros trabalhos que queiram utilizar de tais dados para automatizar futuramente o processo de tradução no domínio de *esports* com maior precisão. Uma aplicação prática de tal funcionalidade é vista na *pesquisa Named Entity Recognition and Named Entity on Esports Contents* (Liu; Leng; Wang; Lin, 2020).

A pesquisa conduzida por Liu e sua equipe utilizou de documentações de jogos e *crawlers* para identificar termos e palavras de destaque em notícias relacionadas a *esports*, o que ajudaria não só na coleta de dados multilíngues, mas também em uma tradução mais precisa.

Uma das maiores qualidades da Linguística de Corpus é a sua pluralidade de aplicações por meio de dados e estatísticas. Tomando o presente corpus como base, a pesquisa como um todo é vista como satisfatória no que conseguiu disponibilizar dados para o uso livre para pesquisa em uma área em crescente expansão.

REFERÊNCIAS

- ALMEIDA, G. M. B. ; CORREIA, M. . **Terminologia e corpus: relações, métodos e recursos**. In: Stella E. O. Tagnin; Oto Araújo Vale. (Org.). *Avanços da Lingüística de Corpus no Brasil*. 1 ed. Sao Paulo: Humanitas/FFLCH/USP, 2008, v. 1, p. 63-93.
- ALMUGHRABI, Zakaria. **THE HISTORY OF LEAGUE OF LEGENDS WORLDS**. 2021. Disponível em: <https://www.hotspawn.com/league-of-legends/guides/the-history-of-league-of-legends-worlds>. Acesso em: 25 set. 2023.
- AntConc. 4.2.4. [Tokyo]: Laurence Anthony, 2001. Software. Disponível em: <https://www.laurenceanthony.net/software/antconc/>. Acesso em: 13 nov. 2023.
- APStylebook. **AP style is esports for competitive multiplayer video gaming. Use alternate forms like eSports or e-sports only if part of a formal name**. [S.I.], 5 de abr. 2017. Twitter: @APStylebook. Disponível em: <https://twitter.com/APStylebook/status/849696651111309312>. Acesso em: 16 set. 2023.
- BAKER, Mona. Corpus Linguistics and Translation Studies — Implications and Applications. **Text And Technology**, [S.L.], p. 233, 1993. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/z.64.15bak>.
- BASTARRICA, Maristela Lutz. **Empréstimos linguísticos do inglês: um estudo do léxico do comércio exterior à luz da teoria da variação em terminologia**. UFRGS. Rio de Grande do Sul: Lume, 2009
- BRITISH COUNCIL. **Demandas de Aprendizagem de Inglês no Brasil**. 2013. Disponível em: https://www.britishcouncil.org.br/sites/default/files/demandas_de_aprendizagempesquisacompleta.pdf. Acesso em: 26 set 2023.
- CABRÉ, M. T. **La terminología: teoría, metodología, aplicaciones**. Barcelona, Antártida/Empúries, 1993.
- COLLIS, William. Chapter One: the bits at the beginning. In: COLLIS, William. **The Book of Esports: the definitive guide to competitive video games**. Nova Iorque: Rosettabooks, 2020. Cap. 1, p. 7.
- DAYRELL, Carmen. O USO DE CORPORA PARA O ESTUDO DA TRADUÇÃO: objetivos e pressupostos. **Tradução em Revista**, [S.L.], v. 2015, Especial, p. 87-102, 29 out. 2015. Faculdades Catolicas. <http://dx.doi.org/10.17771/pucrio.tradrev.25346>.
- DIAZ, GENE. **Stopwords Korean (KO)**. 2016. Disponível em: <https://github.com/stopwords-iso/stopwords-ko>. Acesso em: 13 nov. 2023.
- DPLUS KIA. **Canal do YouTube**. Disponível em: https://www.youtube.com/@DplusKIA_LOL. Acesso em: 12 nov. 2023.
- ESPORTS CHARTS. **LoL Tournaments - Full List of League of Legends Esports competitions for 2021**. Disponível em: <https://escharts.com/tournaments/lol?order=peak&year=2021>. Acesso em: 26 set. 2023.

ESPORTS EARNINGS. **League of Legends**. 2023. Valores até 25 de setembro de 2023. Disponível em: <https://www.esportsearnings.com/games/164-league-of-legends>. Acesso em: 25 set. 2023.

Esports Insider. **Esports Around The World: South Korea**. 2023. Disponível em: <https://esportsinsider.com/2023/07/esports-around-the-world-south-korea>. Acesso em: 25 set. 2023.

EUROGAMER. **The OGA**. 1999. Disponível em: <https://archive.is/bF85>. Acesso em: 16 set. 2023.

Excel. 2310. [S. I.]: Microsoft, 1985. Software. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/excel>. Acesso em 13 nov. 2023.

FERREIRA, Jéssica Barbosa. **DOMINANDO O RIFT: UMA LEITURA SOCIOCULTURAL DA INFLUÊNCIA DA COREIA DO SUL NO LEAGUE OF LEGENDS PROFISSIONAL**. 2022. 296 f. Dissertação (Mestrado) - Curso de Educação Física, Universidade Federal do Paraná, Curitiba, 2022. Disponível em: <https://acervodigital.ufpr.br/handle/1884/79865>. Acesso em: 25 set. 2023.

FIGUEIRA, Matthew. **The viewership of the 2015 League of Legends World Championship toppled Riot's expectations**. 2015. Disponível em: <https://www.criticalhit.net/gaming/viewership-2015-league-legends-world-championship-topped-riots-expectations/>. Acesso em: 25 set. 2023.

FOUNDING Interns of League of Legends. [S. I.]: Riot Games, 2016. (9 min.), son., color. Disponível em: <https://www.youtube.com/watch?v=PKbeDeu5Wbk>. Acesso em: 25 set. 2023.

GAMETRICS. **게임순위**. 2023. Disponível em: <http://www.gametrics.com>. Acesso em: 25 set. 2023.

GOETOMO, Filbert. **Esports in Korea: A study on League of Legends team performances on the share price of owning corporations**. 2016. 60 f. Tese - Curso de Bachelor Of Arts, Economics, Claremont McKenna College, Claremont, 2016. Disponível em: https://scholarship.claremont.edu/cmc_theses/1356/. Acesso em: 25 set. 2023.

GOSLIN. **More than 80 million people watched the Worlds 2017 semifinals**. 2017. Disponível em: <https://www.rifthermal.com/lol-worlds/2017/12/19/16797364/league-of-legends-worlds-viewers-statistics>. Acesso em: 25 set. 2023.

HILTSCHER, Julia. A Short History of eSports. In: HILTSCHER, Julia *et al.* **ESports Yearbook 2013/14**. Norderstedt: Books On Demand GmbH, 2015. p. 9-14.

HOWELL, Leo. **League of Legends hosts 14.7 million concurrent viewers during Worlds**. 2016. Disponível em: https://www.espn.com/esports/story/_/id/18221739/2016-league-legends-world-championship-numbers. Acesso em: 25 set. 2023.

Instituto Nacional Sul-Coreano de Língua Coreana. **Translation Guidelines for Building a Korean-Foreign Language Parallel Corpus**. [S. I.]: Hawoo Publishing, 2023. 348 p. ISBN 979-11-6748-090-3 93710.

JIN, Dal Yong. Historiography of Korean Esports: perspectives on spectatorship. **International Journal Of Communication**. Canadá, p. 3727-3745. 2020. Disponível em: <https://ijoc.org/index.php/ijoc/article/view/13795>. Acesso em: 24 set. 2023.

JOHANSSON, S. Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies. John Benjamins Publishing Company, 2007

KAMBEROVIC, Rijad. **League of Legends Player Count: These Are The Stats**. 2023. Disponível em: <https://riftfeed.gg/more/player-count-league-of-legends#:~:text=In%202022%20League%20of%20Legends,What%20is%20this%3F>. Acesso em: 25 set. 2023.

KIM, Sung-Young. Digital development in Korea: Building an Information Society - by Myung Oh and James F. Larson. **Journal Of Communication**, [S.L.], v. 64, n. 2, p. E5-E9, abr. 2014. Oxford University Press (OUP). <http://dx.doi.org/10.1111/jcom.12081>. Disponível em: <https://www.researchgate.net/publication/261603336>. Acesso em: 20 set. 2023.

LIU, Ziyu; LENG, Yifan; WANG, Meiqi; LIN, Congzhu. Named Entity Recognition and Named Entity Linking on Esports Contents. **Proceedings Of The 2020 Federated Conference On Computer Science And Information Systems**, [S.L.], v. 21, p. 189-192, 26 set. 2020. IEEE. <http://dx.doi.org/10.15439/2020f24>.

LOLESPORTS STAFF. **2018 EVENTS BY THE NUMBERS**. 2018. Disponível em: <https://nexus.leagueoflegends.com/en-us/2018/12/2018-events-by-the-numbers/>. Acesso em: 25 set. 2023.

MCCORMICK, Rich. **'League of Legends' eSports finals watched by 32 million people**. 2013. Disponível em: <https://www.theverge.com/2013/11/19/5123724/league-of-legends-world-championship-32-million-viewers>. Acesso em: 25 set. 2023.

MITCHELL, R. League of Legends championship draws nearly 1.7 million viewers. Disponível em: <https://www.engadget.com/2011-06-22-league-of-legendschampionship-draws-nearly-1-7-million-viewers.html>. Acesso em: 25 set. 2023.

MOZUR, Paul. **For South Korea, E-Sports Is National Pastime**. 2014. Disponível em: <https://archive.is/MEAI5>. Acesso em: 24 set. 2023.

NAVER. **Naver Korean-English Dictionary**. 2023. Disponível em: [https://korean.dict.naver.com/koendict/#/search?query=배\] 췌](https://korean.dict.naver.com/koendict/#/search?query=배] 췌). Acesso em: 14 nov. 2023.

NINTENDO POWER FLASH. Canadá: Mattel, v. 9, 1989. Sazonal. Volume 5 (Summer Fall 1989). Disponível em: https://ia800207.us.archive.org/34/items/Nintendo_Power_Flash_Canada_Issue_05_Summer-Fall_1989/Nintendo_Power_Flash_Canada_Issue_05_Summer-Fall_1989.pdf. Acesso em: 20 set. 2023.

NLTK. 3.8.1. [S. I.]: NLTK Project, 2005. Software. Disponível em: <https://www.nltk.org>. Acesso em: 13 nov. 2023.

OH, Myung; LARSON, James F.. Digital Development in Korea: building an information society. **Journal Of Communication**, [S.L.], v. 1, n. 1, p. 1-272, 14 mar. 2011. Routledge. <http://dx.doi.org/10.4324/9780203829127>.

OLOHAN, Maeve. **Introducing Corpora in Translation Studies**. Londres: Routledge, 2004. 232 p. ISBN: 9780203640005.

OP.GG. **Leaderboards**. Disponível em: <https://www.op.gg/leaderboards/tier?region=kr>. Acesso em: 5 jul. 2023.

PESQUISA GAME BRASIL. **No Brasil, público de games corresponde a 74,5% da população**. 2022. Disponível em: <https://www.pesquisagamebrasil.com.br/pt/meiomensagem-no-brasil-publico-de-games-corresponde-a-745-da-populacao/>. Acesso em: 25 set. 2023.

REA, Stephen C.. Crafting Stars: south korean e-sports and the emergence of a digital gaming culture. **Education About Asia: Sports, Culture, and Asia**. [S.I.], p. 22-27. Outono 2016. Disponível em: <https://www.asianstudies.org/publications/ea/archives/crafting-stars-south-korean-e-sports-and-the-emergence-of-a-digital-gaming-culture>. Acesso em: 20 set. 2023.

ROMER, Rafael. **ESports, esports ou e-sports: Qual a grafia correta dos esportes eletrônicos?**: um debate tão antigo quanto a própria nomenclatura. Um debate tão antigo quanto a própria nomenclatura. 2019. Disponível em: <https://www.theenemy.com.br/esports/esports-esports-ou-e-sports-qual-a-grafia-correta-dos-esportes-eletronicos>. Acesso em: 16 set. 2023.

Runas LOL. **Servidores LoL – Tudo o que você precisa saber**. Disponível em: <https://www.runas.lol/blog/servidores-lol/>. Acesso em: 25 set. 2023.

SARDINHA, Tony Berber. Lingüística de Corpus: histórico e problemática. **Delta: Documentação de Estudos em Lingüística Teórica e Aplicada**, [S.L.], v. 16, n. 2, p. 323-367, 2000. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0102-44502000000200005>.

SARDINHA, Tony Berber. **Lingüística de Corpus**. Barueri, SP: Manole, 2004.

SCHWARTZ, Nick. **More people watch eSports than watch the World Series or NBA Finals**. 2014. Disponível em: <https://ftw.usatoday.com/2014/05/league-of-legends-popularity-world-series-nba>. Acesso em: 25 set. 2023.

T1. **Canal do YouTube**. Disponível em: <https://www.youtube.com/@SKTT1>. Acesso em: 12 nov. 2023.

United Nations Population Fund (UNPF). **World Population Dashboard Korea, Republic of**. 2023. Disponível em: <https://www.unfpa.org/data/world-population/KR>. Acesso em: 5 jul. 2023.

VideoSubFinder. 6.10. [S. I.]: Simeon Kosnitsky, 2007. Software. Disponível em: <https://sourceforge.net/projects/videosubfinder/>. Acesso em: 5 nov. 2023.

WAGNER, M. G. **On the Scientific Relevance of eSports.** (H. R. Arabnia, Ed.) International Conference on Internet Computing. **Anais...**CSREA Press, 2 jan. 2007. Disponível em: <http://dblp.uni-trier.de/db/conf/ic/icom2006.html#Wagner06>. Acesso em: 13 set. 2023.

WELO, Eirik. **Encyclopedia of Ancient Greek Language and Linguistics.** [S. I]: Brill, 2013. General Editor: Georgios K. Giannakis. Disponível em: http://dx.doi.org/10.1163/2214-448X_eagll_COM_00000254. Acesso em: 13 nov. 2023.

YU, Jason. **How Korea Embraced eSports and Haven't Looked Back.** 2015. Disponível em: <https://sea.ign.com/esports/92089/feature/how-korea-embraced-esports-and-havent-looked-back>. Acesso em: 24 set. 2023.