UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

Simone Rodrigues Peron-Corrêa

A SEMÂNTICA DE FRAMES NA CONSTITUIÇÃO DE DICIONÁRIOS TEMÁTICOS MULTILÍNGUES PARA USUÁRIOS NÃO-ESPECIALISTAS: INTERFACE, INTERAÇÃO E AVALIAÇÃO

Juiz de Fora
2019

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

Simone Rodrigues Peron-Corrêa

A SEMÂNTICA DE FRAMES NA CONSTITUIÇÃO DE DICIONÁRIOS TEMÁTICOS MULTILÍNGUES PARA USUÁRIOS NÃO-ESPECIALISTAS: INTERFACE, INTERAÇÃO E AVALIAÇÃO

Tese de Doutorado apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, como parte dos requisitos necessários à obtenção do título de Doutora em Linguística.

Orientador: Professor Doutor Tiago Timponi Torrent

Juiz de Fora
Maio de 2019

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Peron-Corrêa, Simone Rodrigues.
A Semântica de Frames na Constituição de Dicionários Temáticos Multilíngues para Usuários Não-Especialistas : Interface, Interação e Avaliação / Simone Rodrigues Peron-Corrêa. -- 2019.
191 p.

Orientador: Tiago Timponi Torrent
Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2019.

1. Dicionários Temáticos Multilíngues. 2. Diciopédia. 3. FrameNet Brasil. 4. Semântica de Frames. I. Torrent, Tiago Timponi, orient. II. Título.

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

FACULDADE DE LETRAS

PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

Simone Rodrigues Peron-Corrêa

A SEMÂNTICA DE FRAMES NA CONSTITUIÇÃO DE DICIONÁRIOS TEMÁTICOS MULTILÍNGUES PARA USUÁRIOS NÃO-ESPECIALISTAS: INTERFACE, INTERAÇÃO E AVALIAÇÃO

Tese de Doutorado submetido ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, como parte dos requisitos necessários à obtenção do título de Doutor em Linguística.

Aprovada por:

Presidente, Prof. Dr. Tiago Timponi Torrent – UFJF

p/ Profa. Dra. Pilar Araúz–U. Granada

p/ Profa. Dra. Flávia Cristina Cruz Lamberti Arraes – UnB

Profa. Dra. Ana Cláudia Peters Salgado – UFJF

Prof. Dr. Ely Edison da Silva Matos – UFJF

31 de Maio de 2019

Juiz de Fora

“ Combati o bom combate, acabei a carreira, guardei a fé.”
2 Timóteo 4:7

Dedico este trabalho aos meus pais, Braz (in memorian) e Wanda, pelo amor incondicional e por sempre terem investido e acreditado nos meus sonhos e ideais. Ao meu esposo Bruno, por ser meu companheiro de todas as horas e melhor amigo. E à minha filha Melissa, que veio coroar esta etapa, trazendo ainda mais alegria e realização.

AGRADECIMENTOS

A Deus, por seu infinito amor, fidelidade e cuidado para com a minha vida! Por me possibilitar realizar este sonho e por ter me proporcionado vivenciar experiências incríveis!

Ao meu orientador prof. Dr. Tiago Torrent, por seu comprometimento, competência, presteza, incentivo e por ter sido tão presente e atencioso em todos os momentos. Foi um grande privilégio poder ter sido sua orientanda nesses oito anos de pós-graduação (especialização, mestrado e doutorado). Aprendi muito e serei eternamente grata por todo ensinamento, apoio e incentivo para o meu crescimento intelectual, acadêmico e profissional! Muitíssimo obrigada pela parceria!

Aos demais professores do Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora, pelo conhecimento e pelas experiências que foram fundamentais em meu aprendizado!

À minha mãe, por seus ensinamentos e por ser tão presente em minha vida! Por ter acreditado nos meus sonhos e por ter me possibilitado chegar até aqui!

Ao meu amado esposo Bruno, por ser companheiro e um grande amigo de todas as horas. Por compreender minha ausência e por sonhar os meus sonhos comigo!

À minha amada filha, Melissa, que todos os dias me ensina o que é o verdadeiro amor!

Aos meus irmãos, Sandra, Sérgio, Alexsander, Sandriana e Elvis (irmão de coração), meus cunhados, Álvaro e Viviane, e meus sobrinhos, Cristhynne, Lívia e Paulo Sérgio, por serem pilares tão fundamentais em minha vida!

Aos meus sogros, Luiza e Itamar, por me tratarem como uma filha e por me auxiliarem tão imensamente nesta jornada.

Aos meus amigos, Eliezer Isidoro, Raquel Rohr e Renata Lima por serem meus incentivadores de longa data.

Aos amigos de profissão das escolas Apogeu, Balão Vermelho Alicerce e Instituto Federal do Sudeste de Minas Gerais, assim como aos alunos de cada uma dessas instituições, por terem sido companheiros de trajetória e participantes deste processo na minha vida.

Aos integrantes da FrameNet Brasil, por compartilharem seus conhecimentos, por serem verdadeiramente uma equipe que trabalha em prol de um projeto maior. Em especial, aos meus amigos Júlia Gonçalves e Alexandre Diniz por terem contribuído ricamente com conversas, diálogos, traduções, artigos acadêmicos, capítulos de livros e muito afeto nesta caminhada. Ao professor Ely Matos, por sua prontidão e suas contribuições na leitura de minha qualificação. Ao Marcelo Viridiano e ao Arthur Lorenzi, pelo trabalho em conjunto na proposição de ajustes e alterações na interface do aplicativo m.knob.

À equipe da plataforma Appsee, que disponibilizou gratuitamente o aplicativo que usei para análise nesta tese.

À Coordenação e à Secretaria do Programa de Pós-Graduação em Linguística da UFJF, pela presteza em tudo, inclusive, pelos apoios financeiros concernentes aos congressos dos quais participei!

Aos membros titulares e suplentes da banca examinadora desta tese, pela disponibilidade em ler e propor contribuições para este trabalho.

Às fundações de apoio a pesquisa, CAPES, por ter me concedido bolsa no início do doutorado, e FAPEMIG, que posteriormente, deu continuidade. Esse investimento foi fundamental para que eu concluísse esta tese.

Enfim, agradeço a todos os familiares, amigos e colegas que acreditaram e torceram por mim nesta trajetória!

RESUMO

Esta tese discute como as teorias linguísticas influenciam o desenvolvimento de recursos lexicográficos multilíngues e moldam a experiência do usuário com esses recursos. Para tanto, nosso objeto de estudo e análise foi a aplicação computacional (mknob), que foi desenvolvida e está sendo aperfeiçoada pela equipe da FrameNet Brasil. O aplicativo, um guia de bolso multilíngue, possui três funções principais, sendo (i) um sistema de recomendação baseado em geolocalização e filtros de interesse; (ii) a Diciopédia, que se caracteriza por ser um dicionário que trata dos domínios do Turismo e dos Jogos Olímpicos, e (iii) um tradutor baseado em frames e estruturas qualia. Para atingir o objetivo proposto, analisamos a função Diciopédia, que é um dicionário eletrônico que abrange os domínios específicos do Turismo e dos Esportes em três idiomas: Português do Brasil, Inglês e Espanhol. Ancorada nos aportes teóricos da Lexicografia (ATKINS, 2008; L'HOMME, 2010,2014), da Semântica de Frames (FILLMORE,1982,1985), da Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995) e da Terminologia baseada em Frames (FABER ET AL. 2005, 2006,2010), analisamos como essas teorias moldam o recurso lexical Diciopédia; assim como verificamos como esse processo afeta a experiência do usuário com o recurso, através de questionários e do uso de ferramentas que monitoraram cada clique do usuário ao usar a Diciopédia. Esta tese tem outros objetivos adjacentes que corroboram para o adensamento da base de dados da FrameNet Brasil, do aperfeiçoamento do aplicativo (mknob) e da sistematização de um processo de desenvolvimento de outras ferramentas para domínio específico. Os resultados obtidos demonstraram a validação da função Diciopédia de modo positivo; apontaram para ajustes necessários, quanto às necessidades dos consulentes; nos auxiliaram no aperfeiçoamento da modelagem computacional já realizada, assim como na implementação de novos dados que se fizeram necessários, para a proposição de um padrão para construção de dicionários eletrônicos fundados em frames na projeção de novos projetos lexicográficos que atendam outros domínios especializados.

Palavras-chave: Dicionários temáticos multilíngues. Diciopédia. FrameNet Brasil. Semântica de Frames.

ABSTRACT

This dissertation discusses how linguistic theories influence the development of multilingual lexicographic resources and shape the user's experience with these resources. For that, our object of study and analysis was the computational application mknob, which was developed and is being improved by the FrameNet Brazil team. The application, a multilingual pocket tour guide, has three main functions, being (i) a recommendation system based on geolocation and filters of interest; (ii) the Dictiopedia, which is characterized by being a dictionary that deals with the domains of Tourism and the Olympic Games, and (iii) a translator based on frames and qualia roles. In order to reach the proposed objective, we analyzed the Dictiopedia function, which is an electronic dictionary that covers the specific domains of Tourism and OlympicSports in three languages: Brazilian Portuguese, English and Spanish. Based on the theoretical contributions of the Lexicography (ATKINS, 2008; L'HOMME, 2010, 2014), Frame Semantics (FILLMORE, 1982, 1985), Generative Lexicon Theory (PUSTEJOVSKY, 1995) and Frame-based Terminology (FABER *et al*., 2005, 2006, 2010), we analyzed how these theories shape the lexical resource Diciopedia;as well as how this process affects user's experience with the resource through questionnaires and the use of tools that monitored each user's click when using the Dictiopedia. This dissertation has other adjacent objectives that corroborate the densification of the FrameNet Brasil database, the improvement of the application (mknob) and the systematization of a process of development of other tools for specific domain. The obtained results demonstrated the validation of the Dictiopedia mode functioned in a positive way. The results also indicated necessary adjustments to the needs of the consultants; helped us to improve the computational modeling already carried out, as well as the implementation of new data that were necessary for the proposition of a standard for building electronic dictionaries based on frames in the projection of new lexicographic projects that attend other specialized domains.

Keywords: Multilingual thematic dictionaries. Dictiopedia. FrameNet Brasil. Frame Semantics.

SUMÁRIO

LISTA DE FIGURAS xi
LISTA DE GRÁFICOS xiv
LISTA DE TABELAS xv
INTRODUÇÃO 17
1 OS ESTUDOS DO LÉXICO E SUAS PRÁTICAS 20
1.1 Lexicografia 20
1.2 Terminologia 47
1.2.1 Terminografia 51
1.2.2 Constituição de Dicionários de Domínio Específico 54
2 BASES DE CONHECIMENTO FUNDADAS EM FRAMES 60
2.1 Da Semântica de Frames à FrameNet 60
2.2 Desenvolvimento de Recursos de Domínio Específico 75
2.2.1 EcoLexicon 76
2.2.2 DicoEnviro 82
2.2.3 DicoInfo 88
2.2.4 Kicktionary 95
2.2.5 Field: Dicionário de Expressões do Futebol 104
2.2.6 Dicionário COPA 2014 FrameNet Brasil 106
2.2.7 Tabela comparativa dos aspectos lexicográficos das ferramentas descritas 113
3 METODOLOGIA 118
3.1 Prospecção de Termos para o Turismo 119
3.2 Análise de Interação Usuário-Interface 123
3.3 Retomada e esquematização do processo lexicográfico 125
4 DESENHO DA FUNÇÃO DICIOPÉDIA DO M.KNOB 127
4.1 Estruturação do Verbete 129
4.2 Ferramentas de Hipertexto 133
4.3 Ferramentas Colaborativas 135
4.4 Análise dos aspectos lexicográficos concernentes ao aplicativo m.knob 137
4.5 Quantitativos de Entradas Lexicais por Língua 138
5 ANÁLISE DE USABILIDADE DA FUNÇÃO DICIOPÉDIA DO MKNOB 141
5.1 Análise do questionário de impressões de uso da função Diciopedia do Aplicativo m.knob 141
5.2 Análise dos dados fornecidos pela plataforma Appsee 155
5.2.1 Estatísticas de Visualização de Telas pelos Usuários 158
5.2.2 Fluxo de Usuários 161
5.2.3 Touch Heat Maps 164
6 PROPOSTA DE NOVA INTERFACE PARA O VERBETE DA DICIOPÉDIA 173
7 CONCLUSÕES 176
REFERÊNCIAS BIBLIOGRÁFICAS 180
ANEXO I 191

LISTA DE FIGURAS

Figura 1: Exemplo de homônimos homógrafos homófonos no Dicionário Houaiss 24
Figura 2: Definição de homônimos homógrafos não homófonos no Dicionário Houaiss 24
Figura 3: Definição de homônimos homógrafos não homófonos no Dicionário Houaiss 24
Figura 4: Definição da entrada da palavra homógrafa “jaca” no Dicionário eletrônico Michaelis 25
Figura 5: Definição da entrada da palavra homógrafa “bordo” no Dicionário eletrônico Michaelis 27
Figura 6: Definição da entrada da palavra homógrafa “coco” no Dicionário eletrônico Michaelis 29
Figura 7: Definição da entrada do verbo “dar” no Dicionário Houaiss. 29
Figura 8: Definição da entrada da palavra polissêmica “labaça”, no Dicionário eletrônico Michaelis 30
Figura 9: Definição da entrada da palavra polissêmica “cravo”, extraída do Dicionário eletrônico Michaelis 31
Figura 10: Apresentação do verbete desmatar no DiCoEnviro 36
Figura 11: Gráfico que representa a tipologia das decisões de design de dicionário de comércio Fonte: ATKINS, 2008, p.36 41
Figura 12:Comparativo entre Lexicografia e Terminologia 53
Figura 13: Frame Fazer_turismo 64
Figura 14: A modelagem do Cenário do Turismo na FrameNet Brasil 66
Figura 15: Anotação lexicográfica da UL “apreciar” do Frame Fazer_turismo 67
Figura 16: A interface da tabela de entrada para o frame Touring 70
Figura 17: Resultado dos papeis qualia para a UL “mapa.n”. 72
Figura 18: Relações qualia na FrameNet Brasil 73
Figura 19: Definição do frame Turismo 74
Figura 20: Definição do frame Pessoas_por_atividade_de_lazer 75
Figura 21: Tela inicial do dicionário EcoLexicon 76
Figura 22: Conceitos básicos do frame conceitual MEIO AMBIENTAL 79
Figura 23: Mapa conceitual do frame de Marulho 81
Figura 24: Índice alfabético e lista de termos de consulta no DicoEnviro. 83
Figura 25: Critérios de consulta no DicoEnviro. 84
Figura 26: Anotação da UL *reduzir* no DicoEnviro 85
Figura 27: Definição do frame *Charger* (processo de carregamento de uma bateria), no DicoEnviro 86
Figura 28: Relações Lexicais no DicoEnviro. 87
Figura 29: Relações Lexicais, Paradigmáticas e Sintagmáticas no DicoEnviro. 88
Figura 30: Modo de pesquisa por ordem alfabética no DicoInfo. 90
Figura 31: Modo de pesquisa por digitação no DicoInfo. 90
Figura 32: Resultado de pesquisa do termo “web” no DicoInfo. 90
Figura 33:Resultado de pesquisa do termo “écran” no DicoInfo. 91
Figura 34: Entrada do termo “web” no DicoInfo. 92
Figura 35: Entrada com acepções marcadas por letras diferentes no DicoInfo. 92

Figura 36: Exemplo de um termo associado a uma imagem no DicoInfo. 93
Figura 37: Definição e estrutura atuacional no DicoInfo. 93
Figura 38: Ilustração da estrutura actancial no DicoInfo. 93
Figura 39: Ilustração da estrutura atuancial no DicoInfo. 94
Figura 40: Ilustrações de contexto e contexto anotado no DicoInfo. 94
Figura 41: Ilustração das relações paradigmáticas para o termo “navegador” no DicoInfo. 95
Figura 42: Descrição de parte da cena “passe” no Kicktionary. 97
Figura 43: Esquema da estrutura do banco de dados do Kicktionary. 99
Figura 44: Tela inicial do Kicktionary. 100
Figura 45: Tela de busca através da UL no Kicktionary. 100
Figura 46: Tela de resultado pela busca da UL “attacker” no Kicktionary. 101
Figura 47: Tela de busca através das cenas no Kicktionary. 102
Figura 48: Tela de busca através das relações de hierarquia conceptual no Kicktionary. 102
Figura 49: Resultado da busca nos exemplos de fala / narração no Kicktionary. 103
Figura 50: Textos paralelos alinhados com ULs disponíveis para consulta no Kicktionary. 103
Figura 51: Tela inicial do dicionário de expressões futebolísticas Field. 104
Figura 52: Busca por meio de palavras no dicionário de expressões futebolísticas Field. 105
Figura 53: Informação encontrada no cenário “Chute” no dicionário de expressões futebolísticas Field. 106
Figura 54: Interface inicial do dicionário COPA 2014 108
Figura 55: Tutorial do dicionário COPA 2014 109
Figura 56: Possibilidades de busca no dicionário COPA 2014 109
Figura 57: Resultado da consulta por “buscar palavra” no dicionário COPA 2014 110
Figura 58: Tela de consulta por “digitar a frase” no dicionário COPA 2014 111
Figura 59: Resultado da consulta por “ver significado” no dicionário COPA 2014. 111
Figura 60: Resultado da consulta por “explorar a rede” no dicionário COPA 2014. 112
Figura 61: Resultado da consulta por “explorar a rede” Atividades_do_turista no dicionário COPA 2014. 112
Figura 62: Corpora m.knob domínio do Turismo 120
Figura 63: Tela inicial do aplicativo 128
Figura 64: Resultado de busca do termo “check-in” 129
Figura 65: Resultado de busca da UL “feijoada” 130
Figura 66: Resultado de busca da UL “parque” no mknob. 132
Figura 67: Resultado de busca da UL “feijoada.n” 134
Figura 68: Resultado de acesso às ferramentas de hipertexto. 134
Figura 69: Tela de termo não encontrado 135
Figura 70: Resultado de busca do termo “cavalo” na Diciopédia 146
Figura 71: Resultado a respeito de dados o informações sugeridas pelos usuários para a Diciopédia. 155
Figura 72: Resultado do percentual de interação do usuário com as telas do experimento. 158
Figura 73: Resultado do percentual das telas mais visualizadas da Diciopédia. 159

Figura 74: Resultado das telas com maior tempo de duração de visita médio da Diciopédia. 160
Figura 75: Resultado do fluxo do usuário para a tela “Main” (principal) da Diciopédia. 161
Figura 76: Resultado do fluxo do usuário para a tela da Diciopédia. 162
Figura 77: Resultado do fluxo do usuário para a tela de busca da Diciopédia..... 163
Figura 78: Resultado do fluxo do usuário para a tela de termos da Diciopédia. 164
Figura 79: Touch heat map da tela main 165
Figura 80: Touch heat map da tela de chegada da Diciopédia 167
Figura 81: Touch heat map da tela de resultado de busca da Diciopédia..... 168
Figura 82: Touch heat map da tela de resultado de busca da Diciopédia..... 170
Figura 83: Touch heat map ampliado concernente à tela de Termos da Diciopédia . 172
Figura 84: Proposta de design para a tela do verbete da Diciopédia..... 173

LISTA DE GRÁFICOS

Gráfico 1: 160 primeiros termos gerados pelo TERMOSTAT para o Português. 139
Gráfico 2: 160 primeiros termos gerados pelo TERMOSTAT para o Espanhol. 139
Gráfico 3: Resultado concernente ao percentual de pessoas que usam dicionários. . 142
Gráfico 4: Resultado concernente ao percentual de tipo de dicionário mais acessado. 142
Gráfico 5: Resultado concernente ao percentual do nível de proficiência em inglês. . 143
Gráfico 6: Resultado concernente ao percentual do nível de proficiência em espanhol. 143
Gráfico 7: Resultado concernente ao percentual do nível de proficiência em espanhol. 145
Gráfico 8: Resultado concernente ao percentual do nível de proficiência em espanhol. 145
Gráfico 9: Resultado concernente ao percentual de termos que foram ou não sugeridos. 147
Gráfico 10: Resultado concernente ao percentual do nível de facilidade em acessar a Diciopédia. 147
Gráfico 11: Resultado concernente ao percentual do nível de utilidade da Diciopédia para o público-alvo turistas. 148
Gráfico 12: Resultado concernente ao percentual de importância das traduções na Diciopédia. 148
Gráfico 13: Resultado concernente ao percentual das avaliações das traduções na Diciopédia. 149
Gráfico 14: Resultado concernente ao nível de relevância dos nomes dos frames nos resultados encontrados na Diciopédia. 150
Gráfico 15: Resultado concernente ao nível de relevância dos hiperlinks presentes na Diciopédia. 150
Gráfico 16: Resultado concernente ao nível de relevância dos hiperlinks presentes na Diciopédia. 151
Gráfico 17: Resultado concernente ao nível de relevância das estruturas qualia presentes na Diciopédia. 152
Gráfico 18: Resultado concernente ao acesso das ULs que compõem o mesmo frame na Diciopédia. 152
Gráfico 19: Resultado concernente à lista de ULs do mesmo frame quanto ampliar o leque de busca na Diciopédia. 153
Gráfico 20: Resultado concernente ao aspecto de relacionar a definição de com uma rede de significado. 153
Gráfico 21: Resultado concernente a avaliação das informações trazidas pela Diciopédia. 154

LISTA DE TABELAS

Tabela 1: Relações qualia Ternaria no banco de dados m.knob 73
Tabela 2: Tabela dos dados anotados do Kicktionary. 96
Tabela 3: Aspectos lexicográficos das ferramentas descritas 115
Tabela 4: Resultado dos termos gerados pelo TERMOSTAT para o domínio do Turismo em Português. 121
Tabela 5: Dados para análise dos termos selecionados pelo TERMOSTAT 122
Tabela 6:Resumo dos aspectos lexicográficos do m.knob 138

INTRODUÇÃO

A pesquisa de doutorado que ora se apresenta está vinculada à rede semântica FrameNet Brasil (FN-Br), sediada na Universidade Federal de Juiz de Fora (MG), sendo caracterizada por ser um desdobramento do projeto em lexicografia computacional FrameNet, desenvolvido no *International Computer Science Institute* – ICSI, em Berkeley, na Califórnia, desde 1997.

Mais especificamente, debruça-se sobre uma aplicação para usuários não especialistas nos domínios do Turismo e dos Esportes: o m.knob – *Multilingual Knowledge Base* – (mknob.com), produzido, inicialmente, para atender às demandas das Olimpíadas de 2016, sediadas no Rio de Janeiro. É importante ressaltar que, mesmo depois do evento, o aplicativo pode ser usado como um guia turístico multilíngue. Tal recurso é uma aplicação computacional legível tanto por máquina quanto por humanos, cuja função primária é disponibilizar para o turista, seja ele brasileiro ou estrangeiro, assim como para profissionais da área da linguagem (e.g. tradutores e professores de língua; professores de português como segunda língua), um guia turístico multilíngue (português, inglês e espanhol), que recomenda locais, sugere atividades aos turistas, além de traduzir sentenças (En, Es, Pt) nos domínios do Turismo e dos Jogos Olímpicos. Ao se cadastrar o usuário irá escolher uma das três línguas para acessar o aplicativo, podendo também alterar para acessar os outros idiomas, caso queira mudar o direcionamento da busca, de acordo com a pertinência multilíngue. O tradutor automático está em desenvolvimento e o seu diferencial está em propor traduções com base na Semântica de Frames (FILLMORE,1982, 1985), que elege os padrões de valência das Unidades Lexicais no ranqueamento das escolhas, além da implementação de novas relações ancoradas nas estruturas *qualia* (PUSTEJOVSKY, 1995).

Além das funções de guia de viagens e tradutor, o m.knob apresenta ainda um dicionário eletrônico, também fundado na Semântica de Frames, chamado Diciopédia. Na data do lançamento da versão alpha para os Jogos Olímpicos, a Diciopédia computava 2.316 termos, sendo 1.153 em português, 777 em inglês e 386 em espanhol. Esses números apontavam para a necessidade de melhoria desse recurso, não só de modo a equilibrar o quantitativo de termos, mas, principalmente, no sentido de pensá-lo como produto lexicográfico eletrônico, dadas, de um lado, as potencialidades do ambiente eletrônico e, de outro, a densidade da base de dados da FrameNet Brasil e a necessidade de filtrar os dados exibidos no recurso, conforme o perfil do usuário. Dessa

forma, o desenvolvimento adequado da função Diciopédia do m.knob foi a questão motivadora para esta pesquisa.

Em um estágio inicial, procurando atender a essa necessidade, procedeu-se à investigação de métodos automáticos para a extração de termos a partir de *corpora*. Os resultados dessa primeira etapa, desenvolvida com auxílio do *software* TERMOSTAT (DROUIN, 2003), foram publicados em Peron-Corrêa e Torrent (2017).

Entretanto, o problema de pesquisa inicial ganhou contorno muito mais amplo, dados os problemas encontrados nessa etapa e, em especial, os desdobramentos oriundos dos problemas de interface gerados quando o referido dicionário eletrônico temático multilíngue baseado em frames passou a contar com milhares de entradas, organizadas em uma densa rede de relações semânticas e lexicais. Assim, o problema de pesquisa desta tese pode ser assim formulado:

Quais parâmetros metodológicos devem guiar o desenvolvimento de dicionários eletrônicos temáticos multilíngues baseados em recursos lexicais?

Para responder a essa questão, esta tese se baseia nos aportes teóricos da Lexicografia (ATKINS, 2008; L'HOMME, 2010,2014), da Terminologia baseada em Frames (FABER ET AL., 2005, 2006,2007) e da Semântica de Frames (FILLMORE,1982,1985). Dessa forma, coloca sob investigação a seguinte hipótese:

A relevância da estrutura de dados apresentada ao usuário correlaciona-se diretamente com as possibilidades ofertadas pela teoria linguística que dá sustentação ao recurso lexicográfico, considerando o seu propósito.

Nesse contexto, esta tese tem como objetivo principal:

Fornecer uma resposta quantificável do que é útil e relevante para o usuário de dicionário eletrônico temático multilíngue baseado em frames, a partir da testagem de metodologia desenvolvida para a elaboração de tais dicionários.

Apresenta, ainda, os seguintes objetivos adjacentes:

(i) realizar a prospecção de termos para o domínio do Turismo, de modo a aperfeiçoar o recurso lexical eletrônico já existente, através do uso das ferramentas TERMOSTAT e Word List da plataforma Sketch Engine;

(ii) elaborar uma proposta de estruturação dos verbetes da Diciopédia, assim como o funcionamento das ferramentas de hipertexto e colaborativas presentes no aplicativo, de modo a explorar ao máximo as potencialidades do meio eletrônico na constituição de recursos lexicais;

(iii) efetuar uma avaliação da função Diciopédia, como interface de consulta;

(iv) por fim, ao alcançar os objetivos de (i) a (iii), adensar a base de dados da FrameNet Brasil.

Os resultados obtidos demonstram que a função Diciopédia do aplicativo m.knob foi validada positivamente pelos usuários, o que também nos possibilitou constatar os pontos negativos que deveriam ser ajustados, para atender melhor os consulentes, tendo em vista, que se trata de um produto eletrônico, que pode-se manter atualizado, sem tantas demandas de edições, como ocorrem com os dicionários impressos. Essas informações e análises nos auxiliaram no aperfeiçoamento da modelagem computacional já realizada, na implementação de novos dados que se fizerem necessários, na proposição de um padrão para construção de dicionários eletrônicos fundados em frames, assim como na definição de parâmetros para novos projetos lexicográficos que atendam outros domínios especializados.

Esta tese está organizada da forma que se expõe a seguir. No capítulo 1, realizamos um levantamento teórico sobre os estudos do léxico e suas práticas, tais como a Lexicografia, Terminologia, Terminografia e Constituição de Dicionários de Domínio Específico. No capítulo 2, discursamos sobre as bases de conhecimento fundadas em frames, desde a sua teoria a sua aplicabilidade em diversas ferramentas desenvolvidas pelas framenets multilíngues. No capítulo 3, temos a apresentação da metodologia, principalmente, da extração e validação de termos para o domínio do Turismo, assim como a descrição dos questionários e testes aplicados pelas ferramentas, Google Forms e da plataforma Appsee. No capítulo 4, discutiremos a proposta inicial para a função Diciopédia, uma das funções do m.knob, desenvolvida nesta tese. No capítulo 5, apresentaremos a análise de usabilidade da ferramenta Diciopédia. No capítulo 6, desenvolvemos uma nova proposta de interface para os verbetes da Diciopédia. E, por último, no capítulo 7, teremos nossas considerações finais sobre a pesquisa.

1 OS ESTUDOS DO LÉXICO E SUAS PRÁTICAS

A Lexicologia constitui um ramo da Linguística que estuda o léxico, ou seja, as palavras de uma língua e a sua organização sob diversos pontos de vista, por isso, tende a ser compreendida como uma disciplina difusa mediante a sua complexidade e à multiplicidade de abordagens e facetas que envolvem seu objeto de estudo, tais como a etimologia, a formação de palavras, a criação e importação de palavras, a categorização léxico-gramatical, além das relações com a Semântica. Nesse sentido, segundo Krieger e Finatto (2004), a abordagem lexicológica pode ser entendida como o processo em que as unidades léxicas de um idioma são analisadas e descritas, levando em consideração a dualidade forma / conteúdo.

Para Cabré (1999, p.78), o escopo da Lexicologia consiste na construção de um modelo do componente do léxico que possibilite que os conhecimentos implícitos sobre as palavras e o uso que os falantes fazem delas auxiliem nos mecanismos sistemáticos e adequados de conexão entre o componente lexical e os demais componentes gramaticais, permitindo aos falantes de qualquer língua formar novas unidades, seguindo aspectos estruturais sistemáticos. Em suma, a Lexicologia se ocuparia em descrever as palavras da língua, já que sua finalidade é explicar de forma mais adequada possível o funcionamento lexical do falante (CABRÉ,1999, p.80).

A Lexicologia é caracterizada por ser um campo científico transdisciplinar, visto que o tratamento do léxico é um ponto comum de interesse a outras áreas de estudos, devido à multidimensionalidade constitutiva da palavra e das distintas perspectivas que podem ser lançadas sobre ela. No escopo desta tese, interessam os estudos que abordam o léxico por um viés descritivo. Nesse sentido, as próximas sessões apresentarão, em detalhe, a Lexicografia e a Terminologia.

1.1 Lexicografia

A Lexicografia é o ramo da Linguística que estuda o conjunto dos itens lexicais e a descrição do modo como são utilizados. Segundo Landau (1984), é a arte e o ofício de se fazer dicionários, já para Fernández (2003, p.36):

> Teoria lexicográfica, lexicografia teórica ou metalexicografia são as denominações mais usuais para esse componente teórico da lexicografia, que muitos autores começam a empregar para diferenciar essa dimensão teórica tanto da prática concreta ou confecção de dicionários, como de âmbito disciplinar também afim, mas claramente diferenciado por seus objetivos e métodos, da lexicologia.

Para esclarecermos o conceito dado por Fernández, precisamos compreender a autonomia desta área, que procura postular metodologias, princípios e teorias que abarcam tanto problemas quanto propostas de soluções para as aplicações lexicográficas, objetos da lexicografia prática. Para esta tese, discutiremos os dois vieses da lexicografia, o teórico e o prático, de maneira articulada e, para tanto, iremos definir o que é a lexicografia através do principal produto lexicográfico: o dicionário.

Na concepção de Atkins et al. (1994,p.4), os produtos lexicográficos típicos incluem os variados tipos de dicionários impressos, as concordâncias, os índices, as terminologias, entre outros. No contexto das últimas décadas, dada a pervasividade das ferramentas computacionais no fazer científico, aproximam-se desses produtos métodos e ferramentas computacionais destinados a auxiliar as várias tarefas lexicográficas, incluindo a preparação de provas lexicográficas de muitas fontes, a gravação de informação linguística relevante em bancos de dados, a edição de entradas lexicográficas, bem como a divulgação de produtos.

Para compreender melhor a função de um dicionário precisamos conhecê-lo desde os primórdios. Assim, temos que o termo "dicionário" vem do latim: “*dici*”, "*dicere"* que significa ação de "dizer", e o sufixo "ário" indica lugar. Os dicionários existem desde a antiguidade e foram compilados para atender a demandas práticas: por exemplo, os gregos e os romanos recorriam a eles para sanar dúvidas e esclarecer termos e conceitos. Também eram utilizados por viajantes, que usavam listas de palavras em vários idiomas. Nesse contexto, ainda não eram organizados por ordem alfabética e reuniam algumas definições de conteúdos linguísticos ou literários. Conforme elucida Rey-Debove (1984, p. 64 *apud* FARIAS, 1998, p. 77):

> A tradição lexicográfica de listar palavras não ficou restrita apenas à Antiguidade, quando, por exemplo, gregos e romanos elaboraram os primeiros dicionários, mas inclui também o período medieval, caracterizado pela multiplicação de listas bilíngues. Com a evolução das línguas, os dicionários passaram a incluir explicações e comentários a respeito de cada lexema. Hoje em dia, há uma grande diversidade de dicionários, dentre os quais podemos considerar pelo menos três tipos que, segundo Rey-Debove (1984: 64), podem ser assim classificados: (1) o dicionário linguístico que fornece apenas informações sobre os signos, excluindo a definição; (2) a obra enciclopédica que fornece informações sobre as coisas e inclui a definição e (3) o dicionário de língua que informa sobre os signos e inclui também a definição. Ainda segundo Rey-Debove (Ibid.), esses tipos de dicionários podem também ser divididos em dois grupos: o dicionário geral, que trata de todos os signos duma língua dada ou de todas as coisas duma civilização; e o dicionário especial, que só descreve um setor de uma ou da outra.

Mediante essa subdivisão, ainda assim não podemos afirmar que o dicionário geral dê conta de fazer o registro exaustivo de todas as palavras de um determinado idioma,

tendo em vista que ocorre um dinamismo lexical, levando em consideração o aspecto diacrônico (desde a antiguidade até os neologismos mais recentes). Quem cumpre esta função é o *thesaurus*. Dessa forma, com o intuito de cobrir o máximo possível do léxico, baseia-se no maior índice de frequência de uso de uma palavra ou expressão. Trata-se de considerar um nível mais abrangente, diferindo do dicionário especial, ou de domínio específico, no qual a finalidade é reunir subconjunto do léxico de uma determinada terminologia, como, por exemplo, do turismo, das olimpíadas, entre outros, em que os objetos serão definidos a partir de critérios específicos.

O dicionário é, por excelência, uma ferramenta onde encontramos o léxico de um idioma, sendo caracterizado como o lugar formal para a institucionalização do código normativo de uma língua, ou seja, instrumento essencial para a sociedade de cultura, que o destaca como elemento de autoridade sobre o desígnio das "verdades linguísticas". Neste sentido, Oliveira e Isquerdo (2001), afirmam que:

> Na medida em que o léxico configura-se como a primeira via de acesso a um texto, representa a janela através da qual uma comunidade pode ver o mundo, uma vez que esse nível da língua é o que mais deixa transparecer os valores, as crenças, os hábitos e costumes de uma comunidade, como também, as inovações tecnológicas, transformações socioeconômicas e políticas ocorridas numa sociedade. Em vista disso, o léxico de uma língua conserva uma estreita relação com a história cultural da comunidade. Desse modo, o universo lexical de um grupo sintetiza a sua maneira de ver a realidade e a forma como seus membros estruturam o mundo que os rodeia e designam as diferentes esferas do conhecimento. Assim, na medida em que o léxico recorta realidades do mundo, define, também, fatos da cultura.

Para corroborar este pensamento, Biderman (1984, p.28) afirma que o dicionário é caracterizado por exercer um papel normativo dentro da comunidade de falantes, segundo ela:

> O dicionário é um instrumento cultural que remete tanto à língua como à cultura. O lexicógrafo descreve ambas — língua e cultura — como um todo pancrônico, embora se situe numa perspectiva sincrônica. Um dicionário é constituído de entradas léxicas que ora se reportam a um termo da língua, ora a um elemento da cultura. A entrada tem como seu eixo básico a definição da palavra em epígrafe. Essa definição nada mais é que uma perífrase metalinguística da palavra posta como entrada. Essa é a pedra de toque da tarefa lexicográfica, nem sempre executada adequadamente. Nas culturas ocidentais os dicionários têm-se copiado uns aos outros perpetuando erros e inadequações.

O dicionário, em alguma medida, constitui-se como um dos pilares para a definição que um grupo (o dos lexicógrafos) faz de uma língua, de uma cultura e, por extensão, de outras ferramentas sociais que dependem das duas primeiras. Por isso, de acordo com Atkins (2008, p. 35), deve ser um artefato concebido com cuidado, por veicular informações que necessitam ser definidas com precisão, a partir de competências

linguísticas muito específicas. Assim, os lexicógrafos precisam ter uma compreensão aprofundada de conceitos linguísticos, tais como subcategorização, colocações, fraseologia, polissemia, homonímia ou relações semânticas lexicais, entre outros. Grande parte das decisões dos lexicógrafos é pautada em decisões linguísticas, portanto, tal teoria é de extrema importância para a lexicografia teórica.

Conforme Zampolli (1994, p.9), os linguistas podem auxiliar os lexicógrafos na produção de dicionários, levando-se em especial consideração alguns aspectos como: a detecção de regularidades em variações de significado; a correlação das propriedades sintáticas e semânticas; a adoção de um modelo semântico comum na definição de significados das palavras que pertencem à mesma classe conceitual; a indicação explícita, nos dicionários bilíngues, das propriedades linguísticas da língua de origem referentes aos equivalentes de tradução, entre outros.

Para ilustrar a relevância das decisões linguísticas na produção lexicográfica, faremos uma breve comparação entre dois dicionários reconhecidos, sendo um impresso e outro online, com o objetivo de verificar como a homonímia e a polissemia são tratadas em ambos. Elegemos estes aspectos linguísticos pelos seguintes fatores: (i) a forma como estes conceitos serão abordados influenciará na estrutura da descrição de qualquer porção do léxico, ou seja, será decisiva na configuração de um dicionário; (ii) apresentam possibilidades de tratamento para as ambiguidades lexicais, demonstrando, portanto, a influência das escolhas da teoria linguística sobre a organização do dicionário.

No Dicionário Houaiss da Língua Portuguesa, um dos mais tradicionais e conhecidos para o nosso idioma, na parte em que se apresentam os detalhamentos dos verbetes e outras informações técnicas, o usuário encontra, de maneira subdivida, as diversas constituições de homônimos. Apresentaremos cada uma delas, evidenciando as particularidades e restrições às quais estão relacionadas.

No referido dicionário, são assim notados os homônimos homógrafos homófonos:

> Um algarismo alceado à esquerda da unidade léxica que se define é usado nos casos de grupos desses vocábulos que se escrevem e pronunciam da mesma maneira, mas têm origens etimológicas diferentes. A sua ordem de numeração de entrada está ligada à data em que entraram no português, quando esta é conhecida. (HOUAISS & VILLAR, 2009, p. XIV)

Na Figura 1, temos as duas entradas para “jaca”, que exemplificam a definição acima:

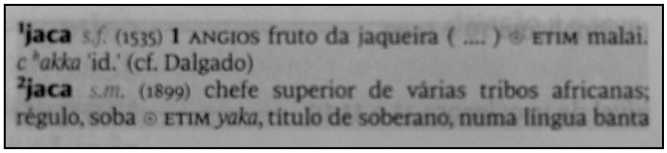


Figura 1: Exemplo de homônimos homógrafos homófonos no Dicionário Houaiss

Fonte: HOUAISS & VILLAR, 2009, p. XIV

No caso da Figura1, o fator decisivo para a constituição da ordem da numeração de entrada no dicionário está relacionado à data em que a palavra surgiu no Português. Neste exemplo, as formas de palavra “jaca” possuem a mesma pronúncia, por isso recebem uma numeração alceada. Entretanto, apresentam suas origens etimológicas diferentes, assim como seus significados.

Em segundo lugar, a chave do dicionário aborda os homônimos homógrafos não homófonos, exemplificados na Figura 2.

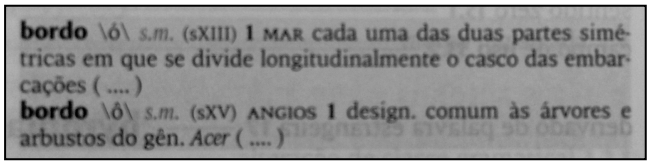


Figura 2: Definição de homônimos homógrafos não homófonos no Dicionário Houaiss

Fonte: HOUAISS & VILLAR,2009, p. XIV

Nesse caso, o que define a ordem na entrada do dicionário é a ortoépia, ou seja, a pronúncia normatizada das palavras, sendo que são apresentadas primeiro as de timbre aberto, conforme se observa pela transcrição fonológica da vogal “o”. Não há uma numeração na entrada, pois a pronúncia do exemplo dado, “bordo” difere em timbre aberto ou fechado.

Os critérios apresentados acima podem ser combinados para casos em que haja homônimos homófonos e não-homófonos, conforme exemplificado na Figura 3.

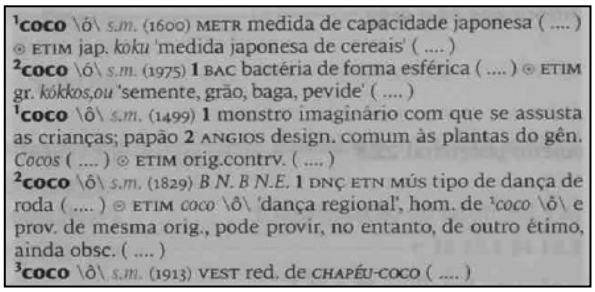


Figura 3: Definição de homônimos homógrafos não homófonos no Dicionário Houaiss

Fonte: HOUAISS & VILLAR,2009, p. XIV

Nesse último caso, podemos encontrar dois ou mais homônimos que apresentam a mesma pronúncia, ou seja, com o mesmo timbre, com o significado diferente, como ocorre no terceiro caso (Figura 3). Nessa situação, a datação de entrada na língua portuguesa continua sendo o referencial para a ordem a ser mencionada no dicionário, como ocorre com o exemplo da palavra “coco”.

Com o intuito de contrastarmos com um dicionário eletrônico, iremos verificar como os homônimos são representados e organizados em tal dicionário. Elegemos o dicionário eletrônico Michaelis Dicionário Brasileiro da Língua Portuguesa, por ser considerado um dos mais usados em nosso país, sendo o acesso gratuito e aberto a todos que queiram realizar a pesquisa, o que o populariza cada vez mais.

No site do dicionário http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/, encontramos os seguintes critérios para a entrada de palavras homônimas: “as palavras homógrafas que possuem etimologias próprias são registradas como entradas diferentes, com número sobrescrito, ou alceado”.

Vejam-se como exemplos as palavras “jaca”, na Figura 4, “bordo”, na Figura 5, e “coco”, na Figura 6.

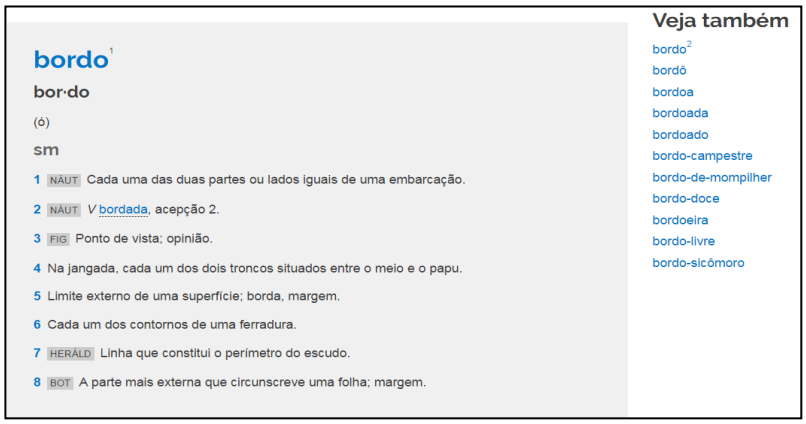


Figura 4: Definição da entrada da palavra homógrafa “jaca” no Dicionário eletrônico Michaelis

Fonte: http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/. Acesso em 24 de setembro de 2016.

As entradas para “jaca”, que representam homônimos homógrafos homófonos, são também alceadas, assim como no dicionário de papel, contudo, no momento da busca da palavra, aparece apenas uma das entradas e, ao lado, constam as outras possibilidades para serem visualizadas, sendo necessário clicar sobre as outras entradas para acessar as demais definições. Ademais, não temos, para o dicionário eletrônico pesquisado, o critério da data de surgimento na língua como ordenador das entradas.

Para o mesmo exemplo de homônimo homógrafo não homófono apresentado pelo Dicionário Houaiss, temos, na Figura 5, a definição encontrada para o termo “bordo”, a qual também segue o mesmo critério de apresentar as demais entradas alceadas e numeradas ao lado. Apesar de seguir o mesmo critério de apresentar, primeiramente, a alternativa com vogal aberta, tal informação não é mostrada ao consulente, uma vez que o dicionário eletrônico em questão não apresenta transcrição fonológica das entradas, ou mesmo recurso multimídia que permita ao usuário ouvir a pronúncia da palavra. Por outro lado, apresentam-se como links as principais expressões com o uso do termo pesquisado.



bordo[1]

bor·do

(ó)

sm

1 NÁUT Cada uma das duas partes ou lados iguais de uma embarcação.

2 NÁUT *V* bordada, acepção 2.

3 FIG Ponto de vista; opinião.

4 Na jangada, cada um dos dois troncos situados entre o meio e o papu.

5 Limite externo de uma superfície; borda, margem.

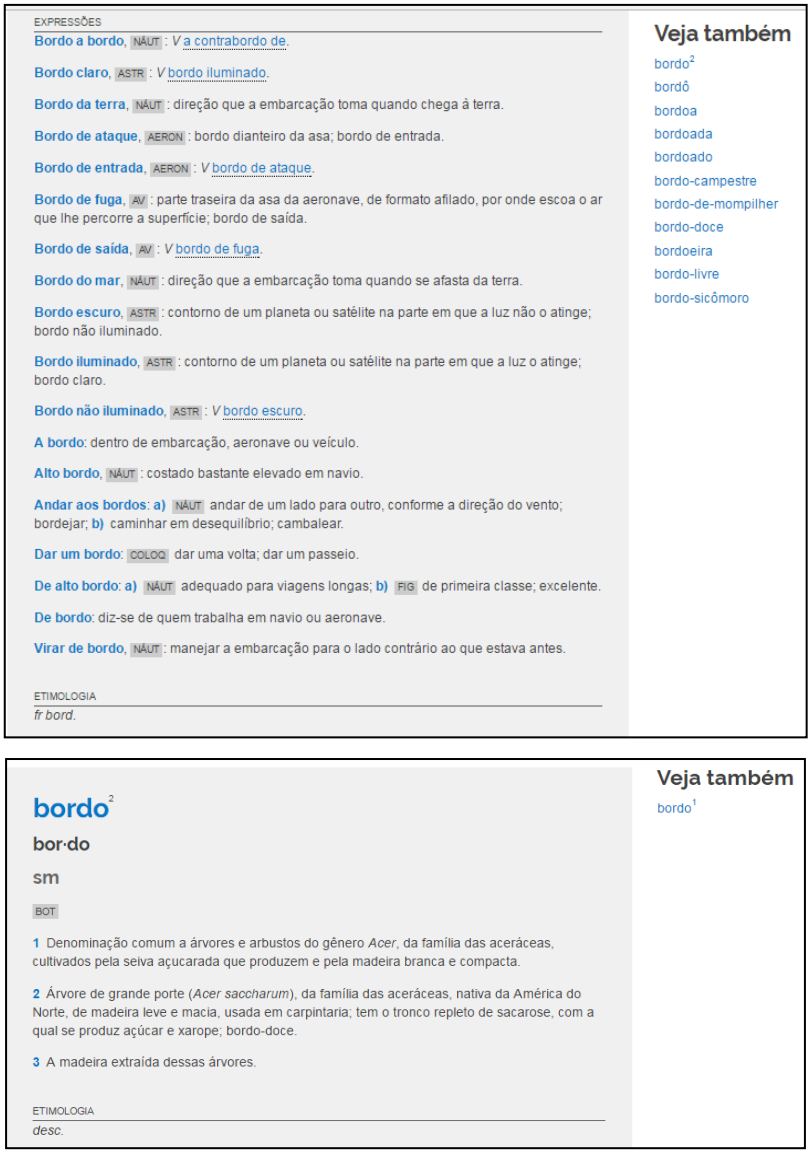
6 Cada um dos contornos de uma ferradura.

7 HERÁLD Linha que constitui o perímetro do escudo.

8 BOT A parte mais externa que circunscreve uma folha; margem.

Veja também

bordo[2]
bordô
bordoa
bordoada
bordoado
bordo-campestre
bordo-de-mompilher
bordo-doce
bordoeira
bordo-livre
bordo-sicômoro



EXPRESSÕES

Bordo a bordo, NÁUT : *V* a contrabordo de.

Bordo claro, ASTR : *V* bordo iluminado.

Bordo da terra, NÁUT : direção que a embarcação toma quando chega à terra.

Bordo de ataque, AERON : bordo dianteiro da asa; bordo de entrada.

Bordo de entrada, AERON : *V* bordo de ataque.

Bordo de fuga, AV : parte traseira da asa da aeronave, de formato afilado, por onde escoa o ar que lhe percorre a superfície; bordo de saída.

Bordo de saída, AV : *V* bordo de fuga.

Bordo do mar, NÁUT : direção que a embarcação toma quando se afasta da terra.

Bordo escuro, ASTR : contorno de um planeta ou satélite na parte em que a luz não o atinge; bordo não iluminado.

Bordo iluminado, ASTR : contorno de um planeta ou satélite na parte em que a luz o atinge; bordo claro.

Bordo não iluminado, ASTR : *V* bordo escuro.

A bordo: dentro de embarcação, aeronave ou veículo.

Alto bordo, NÁUT : costado bastante elevado em navio.

Andar aos bordos: a) NÁUT andar de um lado para outro, conforme a direção do vento; bordejar; **b)** caminhar em desequilíbrio; cambalear.

Dar um bordo: COLOQ dar uma volta; dar um passeio.

De alto bordo: a) NÁUT adequado para viagens longas; **b)** FIG de primeira classe; excelente.

De bordo: diz-se de quem trabalha em navio ou aeronave.

Virar de bordo, NÁUT : manejar a embarcação para o lado contrário ao que estava antes.

ETIMOLOGIA

fr bord.

Veja também

bordo[2]
bordô
bordoa
bordoada
bordoado
bordo-campestre
bordo-de-mompilher
bordo-doce
bordoeira
bordo-livre
bordo-sicômoro

bordo[2]

bor·do

sm

BOT

1 Denominação comum a árvores e arbustos do gênero *Acer*, da família das aceráceas, cultivados pela seiva açucarada que produzem e pela madeira branca e compacta.

2 Árvore de grande porte (*Acer saccharum*), da família das aceráceas, nativa da América do Norte, de madeira leve e macia, usada em carpintaria; tem o tronco repleto de sacarose, com a qual se produz açúcar e xarope; bordo-doce.

3 A madeira extraída dessas árvores.

ETIMOLOGIA

desc.

Veja também

bordo[1]

Figura 5: Definição da entrada da palavra homógrafa “bordo” no Dicionário eletrônico Michaelis

Fonte: http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/. Acesso em 24 de setembro de 2016.

Prosseguindo a pesquisa contrastiva dos mesmos termos nos dicionários, temos na Figura 6 a definição do vocábulo “coco”. Clicamos em cada uma das entradas, para verificarmos cada definição e expressões correspondentes às respectivas definições no dicionário eletrônico Michaelis. Note-se que, além das cinco entradas para “coco”, o dicionário eletrônico também apresenta na seção “Veja também” outras possibilidades ortográficas semelhantes, o que se deve ao fato de, nesta modalidade, a busca por palavras ser feita via digitação e, não, leitura.



coco[1]

co·co

(ó)

sm

1 BACTER Bactéria, geralmente esférica, de dimensão variável.

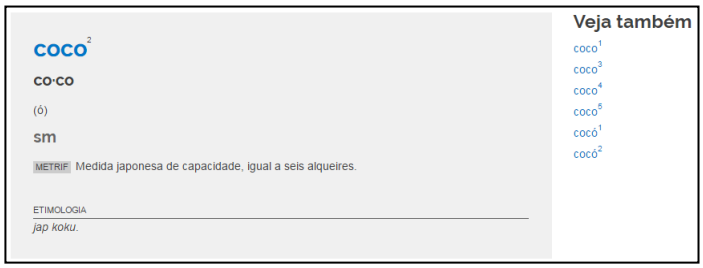
2 BOT *V* coca[4].

ETIMOLOGIA

gr kókkos.

Veja também

coco[2]
coco[3]
coco[4]
coco[5]
cocó[1]
cocó[2]
cocó
coco-amargoso
coco-babão
coco-baboso
cocobacilar



coco[2]

co·co

(ó)

sm

METRIF Medida japonesa de capacidade, igual a seis alqueires.

ETIMOLOGIA

jap koku.

Veja também

coco[1]
coco[3]
coco[4]
coco[5]
cocó[1]
cocó[2]



coco[3]

co·co

sm

1 BOT Denominação comum às plantas do gênero *Cocos*, da família das arecáceas, com uma única espécie.

2 BOT *V* coqueiro[1], acepção 1.

3 BOT Fruto do coqueiro, principalmente do coqueiro-da-baía, usado na culinária brasileira. Apresenta um endocarpo lenhoso, duro, no interior do qual se desenvolve a amêndoa, que possui polpa branca quando o fruto está maduro. Com as fibras do mesocarpo produzem-se capachos, tapetes, cordas etc.; o albume líquido (água de coco) é um refrigerante nutritivo, e o albume sólido, muito rico em óleo, pode ser transformado em leite de coco; ralado, é utilizado na preparação de muitos pratos.

4 BOT Denominação comum a diversas palmeiras com fruto semelhante ao do coqueiro.

5 BOT O fruto dessas palmeiras.

6 Metade da casca desse fruto, empregada para esfregar o pavimento das casas.

7 Vasilha feita do fruto do coqueiro, usada como caneca.

8 COLOQ *V* cabeça, acepção 1.

9 COLOQ Muito dinheiro; dinheirama.

10 COLOQ *V* papão, acepção 1.

11 COLOQ *V* bolinação.

12 *V* cachaça, acepção 1.

Veja também

coco[1]
coco[2]
coco[4]
coco[5]
cocó[1]
cocó[2]

EXPRESSÕES

Coco da praia, REG (N.E.), DANÇA: *V* bambelô.

Coco de desafio, REG (N.E.), DANÇA, MÚS: coco que usa a forma poética para o desafio.

Coco de embolada, REG (N.E.), DANÇA, MÚS: coco que se apresenta através da embolada.

Coco de engenho, REG (N.E.), MÚS: cantiga de trabalho entoada nos engenhos, na época da moagem.

Coco de ganzá, REG (N.E.), BOT: coco ao som do ganzá.

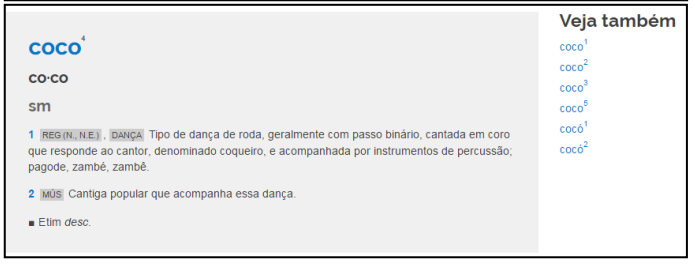
Coco de praia, REG (N.E.), DANÇA: *V* bambelô.

Coco de zambê, DANÇA, MÚS: cantiga e dança de negros com rituais fetichistas, acompanhadas do zambê.

Quebrar coco, REG (N.E.): *V* quebrar catulé.

ETIMOLOGIA

masc de coca[4].



coco[4]

co·co

sm

1 REG (N., N.E.), DANÇA Tipo de dança de roda, geralmente com passo binário, cantada em coro que responde ao cantor, denominado coqueiro, e acompanhada por instrumentos de percussão; pagode, zambé, zambê.

2 MÚS Cantiga popular que acompanha essa dança.

■ Etim *desc.*

Veja também

coco[1]
coco[2]
coco[3]
coco[5]
cocó[1]
cocó[2]

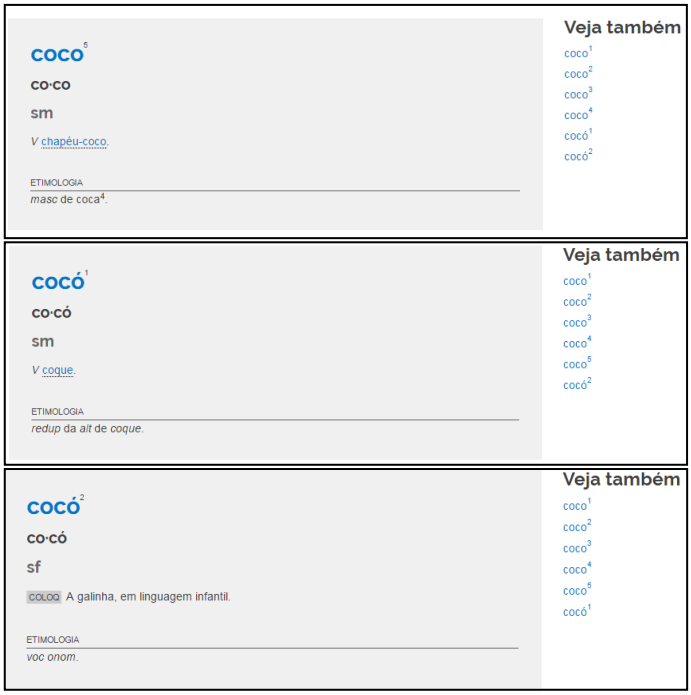


Figura 6: Definição da entrada da palavra homógrafa “coco” no Dicionário eletrônico Michaelis

Fonte: http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/. Acesso em 24 de setembro de 2016.

O segundo aspecto linguístico que iremos exemplificar e comparar é a polissemia, que indica a multiplicidade de significados de uma única palavra. O termo “polis”, no grego significa “muito” e “sema” está relacionado ao significado. Dessa forma, de modo bastante simplificado, consideramos uma palavra polissêmica quando em diferentes contextos, uma palavra pode obter outros significados.

No dicionário Houaiss (HOUAISS & VILLAR, 2009, p. XVI), temos a seguinte regra, exemplificada na Figura 7:

> No caso de o verbete contar com mais de uma acepção, cada uma delas vem sempre antecedida de um numeral em negrito. Quando se trata de acepções restritivas que se seguem a acepções de sentido mais geral, ou no caso de acepções com diversos subitens a ela referentes, o dicionário emprega uma subnumeração (p.ex., **1.1**):

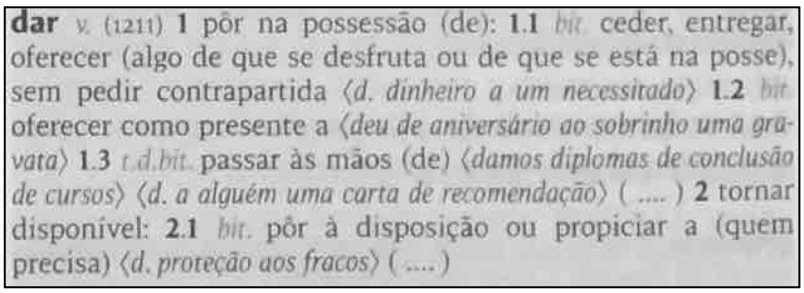


Figura 7: Definição da entrada do verbo “dar” no Dicionário Houaiss.

Fonte: HOUAISS & VILLAR, 2009, p. XVI.

A ordenação das acepções em uma entrada neste dicionário parte das de datação mais recuada para as mais recentes, embora a regra de preferência pela derivação imediata ou a derivação semântica possam alterar tal disposição.

> Idealmente, um primeiro grupo de acepções, organizado segundo a diretiva cronológica, contém todos os sentidos da unidade léxica analisada de nível 'neutro' (as não rubricadas, que fazem parte da língua mais geral ou da culta, literária). A ele se segue um segundo grupo que engloba todas as acepções rubricadas, na ordem de alfabetação de suas rubricas (HIST depois de GEO e antes de MAT, por exemplo). Este segundo grupo não leva em consideração a organização por data. (HOUAISS & VILLAR, 2009, p. XVII)

Ao compararmos com o dicionário online Michaelis, os autores mencionam duas formas de acepção na entrada do dicionário, conforme podemos visualizar na Figura 8:

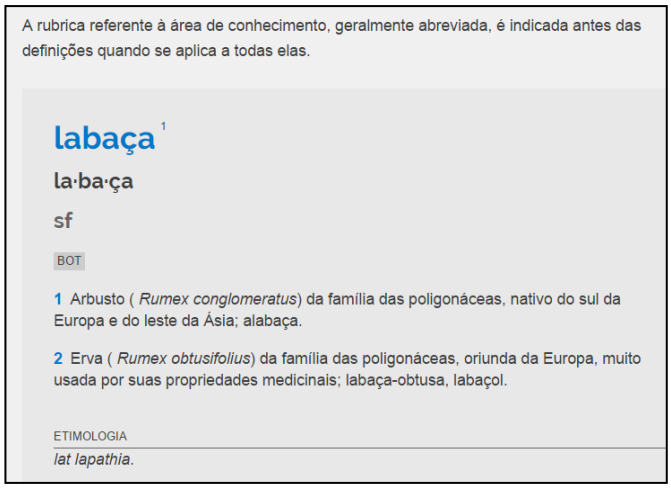
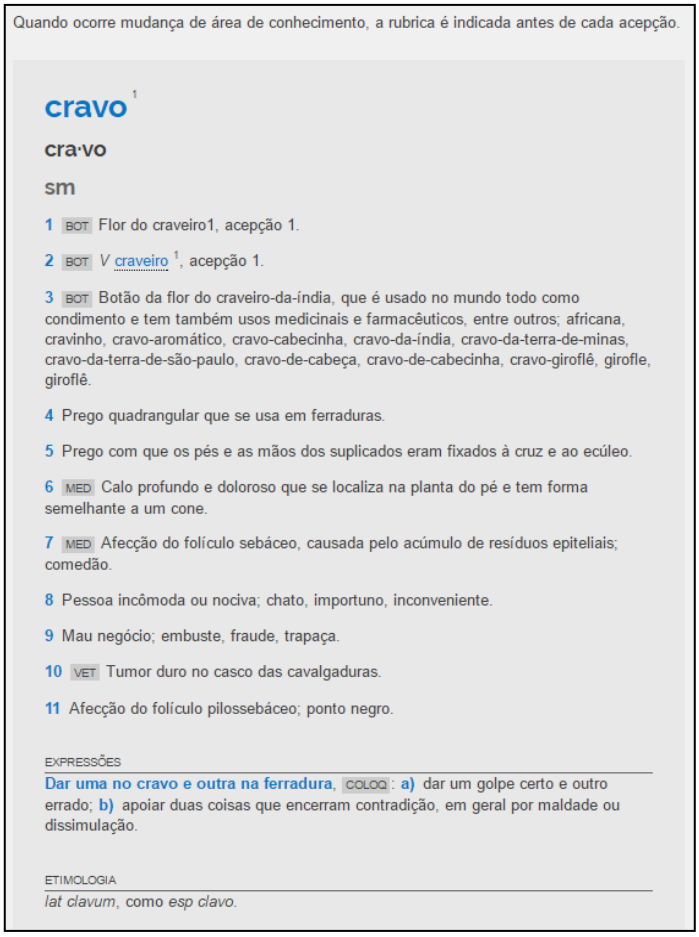


Figura 8: Definição da entrada da palavra polissêmica "labaça", no Dicionário eletrônico Michaelis

Fonte: http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/. Acesso em 10 de fevereiro de 2017.

Como as duas definições da palavra "labaça" são do mesmo campo de conhecimento, a rubrica vem abreviada antes de ambas as definições. Caso sejam, de áreas de conhecimento diferentes, a rubrica será indicada antes de cada acepção, conforme ilustra a Figura 9:



Quando ocorre mudança de área de conhecimento, a rubrica é indicada antes de cada acepção.

cravo [1]

cra·vo

sm

1 BOT Flor do craveiro1, acepção 1.

2 BOT *V* craveiro [1], acepção 1.

3 BOT Botão da flor do craveiro-da-índia, que é usado no mundo todo como condimento e tem também usos medicinais e farmacêuticos, entre outros; africana, cravinho, cravo-aromático, cravo-cabecinha, cravo-da-índia, cravo-da-terra-de-minas, cravo-da-terra-de-são-paulo, cravo-de-cabeça, cravo-de-cabecinha, cravo-giroflê, girofle, giroflê.

4 Prego quadrangular que se usa em ferraduras.

5 Prego com que os pés e as mãos dos suplicados eram fixados à cruz e ao ecúleo.

6 MED Calo profundo e doloroso que se localiza na planta do pé e tem forma semelhante a um cone.

7 MED Afecção do folículo sebáceo, causada pelo acúmulo de resíduos epiteliais; comedão.

8 Pessoa incômoda ou nociva; chato, importuno, inconveniente.

9 Mau negócio; embuste, fraude, trapaça.

10 VET Tumor duro no casco das cavalgaduras.

11 Afecção do folículo pilossebáceo; ponto negro.

EXPRESSÕES

Dar uma no cravo e outra na ferradura, COLOQ: **a)** dar um golpe certo e outro errado; **b)** apoiar duas coisas que encerram contradição, em geral por maldade ou dissimulação.

ETIMOLOGIA

lat clavum, como *esp clavo*.

Figura 9: Definição da entrada da palavra polissêmica “cravo”, extraída do Dicionário eletrônico Michaelis
Fonte: http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/. Acesso em 10 de fevereiro de 2017.

Em ambos os dicionários analisados o que se observa é a existência de um conjunto de decisões teoricamente fundadas que são decisivas para a forma de apresentação do produto final. Apesar de serem tratadas como dadas pelas chaves de leitura dos dicionários, conceitos como “acepção neutra versus acepção rubricada” ou “língua geral versus língua culta ou literária” não são de definição direta e inconteste.

Em vertente complementar à da influência da teoria linguística sobre a confecção de dicionários, mesmo escolhas metodológicas aparentemente ateóricas – por exemplo, a de se listar primeiro a acepção mais antiga da entrada lexical – tem sido questionadas, dada a necessidade de se aprimorar cada vez mais a forma de apresentação das informações no dicionário, de acordo com as demandas do usuário. Nesse sentido, Biderman explica:

> No passado muitos dicionários operavam da seguinte forma: indicavam primeiro o significado original, etimológico, alistando, a seguir, os significados subsequentes até os valores semânticos contemporâneos ao dicionarista. Evidentemente só uma profunda e detalhada pesquisa na documentação histórica da língua e um gigantesco

> banco de dados de todos os períodos do idioma autorizaria os lexicógrafos a estabelecer a escala correta dos significados. Por isso os dicionaristas contemporâneos optaram por um modelo mais exequível e intuitivo, porém mais útil para o consulente do dicionário. Ordenam-se os significados com base na sua maior frequência de uso, ou seja, dos sentidos mais comuns aos menos frequentes. (BIDERMAN,1984, p.36).

Assim, para que o produto lexicográfico atenda as principais necessidades e exigências do usuário, é preciso pressupor, para além de uma base teórica linguística, o conhecimento de como as pessoas realmente usam os dicionários. Para Rigual e Calvi (2014, p.44), antes de compilar um dicionário, seria necessário identificar quem seriam seus potenciais consulentes e suas necessidades lexicográficas específicas. Trata-se de uma atividade dispendiosa e demorada, pois o cerne, principalmente dos dicionários monolíngues, é especificação do significado, ou seja, as definições a que os usuários tendem a recorrer para procurar palavras, a fim de saber mais sobre seus sentidos.

Sob este mesmo enfoque, Atkins (2008, p. 35) propõe que um fator relevante na fase pré-lexicográfica é a definição do grupo de usuários, o que determinará vários fatores, dentre eles, a adequação da linguagem e demais semioses para o público leitor. Isso porque, segundo a autora, há um jogo com regras rígidas, que nem sempre beneficiam o usuário, como, por exemplo, o uso da fonte *Times New Roman* indica um tipo de informação, enquanto o uso do itálico indica outra, entre outros detalhes. Ademais, o critério de seleção do verbete e a forma como é repassado para o usuário independem do fato de estes conhecerem ou não as abreviaturas usadas, como “substantivo”, “verbo” e “transitividade”. Dessa forma, Atkins afirma que quem possui o “trunfo” neste jogo são os compiladores de dicionários, já que:

> O lexicógrafo é quem define as regras do jogo específicas para qualquer dicionário em particular e as define na introdução. Infelizmente, em alguns dicionários, pode parecer que o lexicógrafo pensa: "Uma vez que eu lhes disse quais são as regras, se eles não podem lidar com elas, bem, minha consciência está tranquila”. (ATKINS, 2008, p.37)[1]

Para que o produto lexicográfico seja bem sucedido, ou seja, para que alcance o objetivo pré-estabelecido, é importante que ocorra um ato de comunicação também bem sucedido entre o compilador e o consulente, pois se trata de uma ferramenta para atender as demandas de quem irá consultá-lo, por isso, se não funciona quando usada com inteligência, a culpa é de quem o produziu e não do usuário (ATKINS, 2008, p. 37).

---

[1]The lexicographer gets to define the specific rules of the game for any particular dictionary and sets these out in the introduction. Unfortunately, in some dictionaries, it can appear that the lexicographer thinks, “Now I’ve told them what the rules are. If they can’t manage them, well, my conscience is clear.”

Outros aspectos relacionados à importância do consulente na produção lexicográfica foram discutidos por L'Homme e Cormier (2014), principalmente no que tange aos dicionários digitais e eletrônicos, tais como a facilidade em produzir uma maior quantidade de informações no formato digital, assim como sua versatilidade ao apresentar diferentes partes do conteúdo de uma entrada; o vínculo de palavras dentro de entradas; a extração de listas de palavras relacionadas e os links de um dicionário para outros recursos. Além disso, os usuários podem acessar vários tipos de informações, como dados lexicográficos, dados enciclopédicos, dados terminológicos, dados de corpus, e, em alguns casos, podemos encontrar a integração de dicionários com ferramentas de análise textual (gramaticais, por exemplo).

Segundo eles:

> o formato digital pode [ainda] ser usado para controlar, pelo menos parcialmente, a forma como os usuários interagem com a interface do dicionário e, em seguida, encontrar melhores maneiras de atender às suas necessidades. (L'HOMME e CORMIER, 2014, p.334)[2]

Tal possibilidade é explorada nesta tese, conforme se relatará no capítulo 5. Entretanto, os autores também ressaltam um fator negativo na diversidade de informações que os meios digitais permitem oferecer ao usuário. Muitos desses dados podem ser irrelevantes para o destinatário, sobrecarregando-o de uma grande quantidade de informações desnecessárias, como o que ocorre com os dicionários impressos, o que também acaba sendo reproduzido pelos dicionários digitais. Dessa forma, afirmam que o potencial oferecido pelos computadores não tem sido totalmente explorado para o benefício dos usuários digitais, o que mostra que a abordagem utilizada ainda é a de dicionários impressos com acesso eletrônico.

Lew e Schryver (2014, p. 352) também estão de acordo com L'Homme e Cormier (2014) sobre o tipo de abordagem que ainda vem sendo tomada na lexicografia computacional:

> Dicionários eletrônicos têm apenas parcialmente atingido o seu potencial, porque eles só têm expandido suas capacidades de acesso de maneiras bastante periféricas, apesar de uma avalanche de novas capacidades computacionais. Fundamentalmente, dicionários eletrônicos "pensam" em si mesmos como dicionários impressos a serem oferecidos via acesso eletrônico. Esta é uma visão muito limitadora. (LEW e SCHRYVER, 2014, p.352). [3]

---

[2]the digital format can be used to track, at least partially, the way users interact with the dictionary interface and then find better ways to meet their needs.

[3] Electronic dictionaries have only partially achieved their potential because they have only expanded their access capabilities in fairly minor ways despite an avalanche of new computational capabilities.

Apesar dessas limitações constatadas a respeito do uso dos dicionários eletrônicos, Lew e Schryver (2014, p. 342) também apontam alguns benefícios vinculados ao uso do suporte computacional, com o objetivo de propor que a experiência do usuário seja ainda mais eficiente e mais amigável. Os autores verificaram que o dicionário mais utilizado nos últimos anos é o dicionário eletrônico e que tal fator se deve principalmente ao uso das novas tecnologias, que provocaram uma grande revolução digital, o que interferiu diretamente na lexicografia. Dessa forma, para eles, os dicionários – que sempre foram inerentemente práticos – vieram agora a ser reconhecidos como ainda mais práticos, devido a vários fatores, tais como: (i) atualização com maior frequência em suas edições, ao contrário do que ocorria com os dicionários impressos, os quais dificilmente seriam substituídos pelo usuário; (ii) a opção de ouvir a pronúncia de novas palavras; (iii) a interconectividade com outros recursos (tais como corpora, sites e links); (iv) o incentivo à consulta mais frequente em comparação com dicionários impressos tradicionais, sendo essa consulta ainda mais rápida e de fácil acesso; (v) a facilitação da filtragem nas pesquisas, mediante a flexibilização na organização dos dados; (vi) a transposição, por parte dos dicionários eletrônicos, dos limites formais, por exemplo, da fonologia ou ortografia, uma vez que é possível fazer combinações formais e semânticas para realizarmos a busca; (vii) a possibilidade de realizar a pesquisa de multipalavras, o que os dicionários de papel têm muita dificuldade em realizar.

Fellbaum (2014, p. 378) também descreve os principais ganhos na lexicografia com a revolução digital, corroborando os apontamentos que já foram mencionados anteriormente. Cita a construção de bases de dados lexicais cada vez maiores com representações multifacetadas, flexíveis e ricas de significado das palavras, ultrapassando as restrições no tamanho e no formato dos dicionários impressos; o acesso a corpora de textos eletrônicos, os quais fornecem uma base de dados empírica e sólida para o lexicógrafo consultar e refletir sobre o uso real do falante; a integração das diferentes fontes de pesquisa, independentemente da língua; além da contínua atualização dos dicionários eletrônicos por especialistas, independentemente de novas edições.

Fundamentally, electronic dictionaries ‘think’ of themselves as print dictionaries being offered via electronic access. This is a very limiting vision.

L'Homme (2010, p.361), a partir de sua experiência com os dicionários especializados online (DiCoInfo – Dicionário de informática e termos da Internet e o DiCoEnviro – Dicionário básico do meio ambiente), defende que algumas descrições formais ou semi-formais do léxico, baseadas, no caso específico, em frames, podem ser úteis para codificar a informação lexical em formato digital. Entretanto, cabe ao lexicógrafo considerar em que medida os princípios descritivos utilizados na confecção do dicionário devem estar explícitos para o consulente. Assim, a autora ressalta diferentes estratégias que podem e devem ser utilizadas na seleção de codificações formais conforme o interesse do consulente, colocando-as no *background* ou facilitando sua apresentação com dispositivos de fácil acesso. Embora tais codificações sejam necessárias, lexicógrafos não podem esperar que os usuários decifrem ou se interessem por todas elas (L'HOMME, 2014, p.362).

Para compreendermos melhor esse processo, a autora o descreve de maneira elucidativa. Inicialmente, ressalta que as entradas lexicais no banco de dados continuam sendo estruturadas conforme os princípios de análise definidos pelo lexicógrafo, sendo apenas alterada a forma como a interface apresenta essas informações ao consulente, selecionando as informações mais relevantes da entrada, conforme o perfil do usuário.

A título de exemplo, L'Homme explica a estrutura de actantes utilizada no DiCoEnviro, a qual:

> indica o número de actantes (participantes obrigatórios) de uma unidade lexical e fornece informações complementares, tais como: a) como os actantes são colocados em relação à unidade lexical; b) o papel semântico dos actantes (por exemplo, agente, paciente, fonte); c) termos típicos (aqueles esperados em determinadas posições atuacionais). (L'HOMME, 2014, p. 365).[4]

A primeira decisão pensada para facilitar a interpretação dessa estrutura pelo usuário é a de utilizar o termo típico para representar cada actante, substituindo o nome técnico do actante em si. Tal termo é um item lexical que prototipicamente instancia o actante. Isso facilita a pesquisa para o consulente ao evitar uso de metalinguagem desnecessária. Na apresentação dos termos típicos, não aparece o papel semântico que instanciam, informação esta que compõe o banco de dados. Caso o usuário queira saber, basta passar o *mouse* sobre o termo, e, em uma caixa acima, surge a informação, como

[4]States the number of actants (obligatory participants) of a lexical unit and provides complementary information such as: a) how the actants are placed relative to the lexical unit; b) the semantic role of actants (e.g., Agent, Patient, Source); c) typical terms (the ones expected in given actantial positions).

paciente, agente, etc. Por exemplo, para o verbo *desmatar*, o termo típico para o actante agente seria *homem*, para origem*, floresta* e, para paciente*, árvore*, conforme Figura 10.

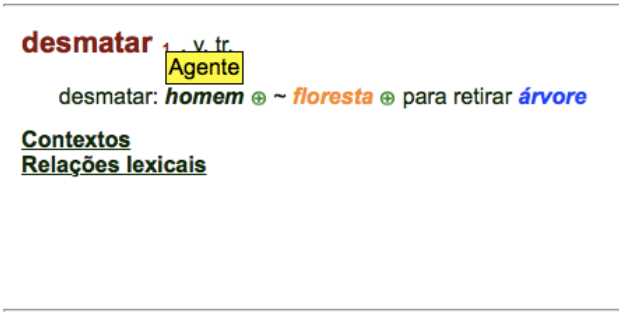


Figura 10: Apresentação do verbete desmatar no DiCoEnviro

Fonte: http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search-enviro.cgi?ui=pt&mode=mot&equi=1&lang=pt&prec=exact&rq=desmatar

Para a seleção dos termos típicos, vários critérios são adotados, como: (i) constar na lista de palavras de temas específicos, como meio ambiente; (ii) aparecer com uma determinada frequência nos corpora analisados; (iii) corresponder a termos genéricos, mas também satisfazer as demandas para compor a lista de termos típicos do domínio.

Dentre as várias vantagens mencionadas pela autora, com a lexicografia eletrônica é possível colocar parte das informações em segundo plano, sendo ainda assim acessível de alguma forma para os usuários que desejam saber mais sobre as palavras que usam.

Dados esses aspectos, podemos verificar que tem sido crescente o aperfeiçoamento de ferramentas lexicográficas, visando a proporcionar maior agilidade e objetividade para o consulente. Nesse sentido, a teoria linguística tem saído apenas do âmbito informacional nas decisões lexicográficas e tem passado a estar presente na interface de consulta de maneira que possa atender aos usuários comuns e até mesmo aos especialistas das diversas áreas.

Mesmo em sua versão impressa, mas mais ainda em formato eletrônico, o dicionário é considerado um texto complexo, que apresenta muito mais do que uma lista de palavras, sendo um objeto semiótico com uma multidimensionalidade de aspectos, que incluem, além da semântica, aspectos sintáticos, como regência preposicional, combinações possíveis ou não, predicação dos verbos, possíveis construções sintático-semânticas, além de informações adicionais, quando necessário. Neste sentido, Krieger (2006) pontua que:

> a obra dicionarística não se resume a uma listagem, mas, como um texto, possui regras próprias de organização. Nessa regularidade organizacional, há projetos específicos

> definidos em conformidade com os fins visados pelo lexicógrafo que imprime suas marcas subjetivas e ideológicas em sua obra (KRIEGER, 2006, p. 141).

Outra concepção que resume bem a essência do dicionário, expandindo o conceito de Krieger para o nível da comunidade de fala[5], foi descrita por Lara (1997, p. 18), que afirma que o dicionário é "uma construção histórica, fruto da reflexão sobre a língua e orientada para a conservação da memória de experiências de sentido valiosas para a comunidade linguística inteira". Assim é que, para compreendermos a produção lexicográfica, precisamos conhecer os aspectos que diferenciam cada tipo de dicionário de acordo com suas especificidades.

Para Atkins (2008, p. 31), os dicionários podem ser analisados sobre dois pontos de vista: forma e conteúdo. O primeiro pode ser entendido como um livro que lista e explica os significados das palavras de uma língua. Mesmo nos dias atuais, com tanta tecnologia e informação, os dicionários eletrônicos são configurados como livros eletrônicos em sua maioria, conforme já se pontuou acima, uma vez que as editoras não investem em dicionários eletrônicos personalizados, por não serem rentáveis. O segundo aspecto, o do conteúdo, aponta para uma classificação em dicionário acadêmico e histórico, que não é tão produzido, devido a custos; ou em dicionário comercial, que é produzido para ser vendido no mercado. Neste sentido, a autora aponta que sua atenção se volta mais para este último, visando a mostrar a importância da lexicografia teórica, já que se trata de produtos comercialmente competitivos.

Grande parte dos dicionários monolíngues modernos faz o uso de dados autênticos e *corpora* eletrônicos com o propósito de incluir informações fiáveis, de acordo com a frequência de ocorrências nas entradas. Isso também faz com que seja possível organizar as entradas polissêmicas com base na frequência de seus significados. Essa abordagem foi lançada pelo projeto Cobuild (SINCLAIR ET AL., 1990), conforme menciona Rundell (2008), no que tange aos avanços tecnológicos que têm contribuído muito para a qualidade da informação, tais como a descrição linguística de como as palavras são usadas e a forma como esta informação é repassada para o usuário. Ainda neste sentido, Zampolli (1994, p.6) afirma que:

---

[5] A comunidade de fala não é definida por nenhum acordo marcado quanto ao uso dos elementos da língua, mas, sobretudo, pela participação em um conjunto de normas compartilhadas. Essas podem ser observadas em tipos claros de comportamentos avaliativos, e pela uniformidade de seus termos abstratos de variação, que são invariáveis com relação aos níveis particulares de uso (LABOV, 1972, p. 120- 121, tradução nossa).

> *Corpora* textuais são agora reconhecidos como a fonte essencial de dados necessários para informar a descrição das línguas naturais em diferentes contextos comunicativos[6].

O objetivo inicial do projeto Cobuild, em 1981, foi o de construir um corpus do inglês atual e usá-lo para produzir um dicionário de inglês para aprendizes, com o intuito de competir com os dicionários Oxford e Longman. As ideias de Sinclair e de sua equipe foram canalizadas em computação léxica na produção do dicionário *Collins Cobuild English Language Dictionary* (CCELD), que teve sua primeira publicação em 1987. Foi considerado um livro de referência por pessoas de toda a parte do mundo, usado por tradutores, aprendizes de língua, professores e também como um texto que produziu um enorme impacto na prática e criação de dicionários comerciais, além do aspecto lexicográfico, na descrição da linguagem em geral.

Ainda que a Linguística de Corpus tenha iniciado de maneira efetiva na década de 1960, com a criação do corpus de Brown e LOB, com 1 milhão de palavras, o aspecto computacional ainda era escasso no final da década de 1970, o que justificava a limitação do uso de corpus por linguistas naquele contexto. Desse modo, o projeto Cobuild, ao utilizar um corpus para criar um dicionário, inovou tecnológica e lexicograficamente em sua experimentação. Segundo Moon (2007, p.163), o projeto Cobuild conseguiu explorar várias vertentes, como produzir um dicionário completamente novo para aprendizes, com base em evidência em corpus, questionando, assim, as convenções lexicográficas de um mercado que já havia sido abalado. O Dicionário Johnson da Língua Inglesa (1755) e o Oxford English Dictionary (1884-1928) basearam-se em evidências de coleções de citações, que, mesmo não sendo consideradas representativas, influenciaram outros dicionários a se basearem nesta mistura de citações e introspecção.

Mesmo o corpus tendo se tornado um recurso lexicográfico comum, com o seu advento vieram também os seus problemas, tais como exigir muito tempo na análise dos dados extraídos, considerado um fator caro para os orçamentos comerciais. Outro aspecto é o crescimento do corpus, o que requeria uma metodologia específica durante os estágios iniciais do projeto Cobuild, em que os lexicógrafos trabalhavam com impressões em papel de linhas de corpus. Um terceiro problema mencionado é que ainda não existia uma compreensão perfeita de como o corpus mapeia o uso do idioma como

---

[6]Textual corpora are now recognized as the essential source of the data needed to inform the description of natural languages in different communicative contexts.

um todo, mostrando a que os usuários estão expostos no cotidiano, sem haver homogeneidade; e por último, constataram que uma abordagem purista dos dados não conduziu, necessariamente, a melhores dicionários.

Uma abordagem central para o projeto Cobuild foi considerar, na análise do léxico, que o sentido e a sintaxe deveriam estar intimamente ligados (MOON, 2007, p. 167). Tal fato sugeriu a necessidade de um realinhamento da importância do sentido e da sintaxe em uma integração mais completa das informações gramaticais nas descrições individuais. Uma série de técnicas foi desenvolvida na edição do CCELD pela equipe Cobuild, entre elas abandonar a teoria da substituição, que consistia em substituir o item ou sentido que está sendo definido, em seu contexto, pela definição sem perda ou mudança de significado. Com isso, foi possível não apenas indicar o contexto, o padrão e o uso, mas evitar qualquer implicação de que o significado exista em isolamento do contexto.

Dentre os aspectos positivos, as definições forneciam uma maneira mais clara de visualizar o ambiente linguístico das palavras tanto do ponto de vista discursivo quanto estrutural. Os significados das palavras comuns, que, geralmente, apresentam múltiplos sentidos, foram diferenciados através da fraseologia e do contexto, uma vez que priorizavam os usos observados em dados, no projeto Cobuild.

> Nem todos os recursos do CCELD foram imitados: alguns foram ignorados, alguns rejeitados de forma definitiva. No entanto, Sinclair, através do CCELD, teve um efeito catalítico sobre a lexicografia, começando com os dicionários de aprendizes monolíngues, espalhando-se para dicionários bilíngues e dicionários monolíngues para falantes nativos adultos ou infantis, na Grã-Bretanha e internacionalmente (MOON, 2007, p. 176)[7]

Mesmo com o pioneirismo no projeto de computação léxica, o projeto Cobuild, infelizmente, negligenciou as necessidades do usuário na descrição do idioma, uma vez que não possibilitou que o dicionário CCELD fosse sinônimo do projeto, sem, entretanto, prejudicar seu papel como marco na produção de dicionários a partir de corpora.

Com o advento da informatização dos corpora, tivemos mudanças fundamentais e irreversíveis na constituição de dicionários. A partir de 1980, lexicógrafos puderam produzir uma descrição muito mais satisfatória dos idiomas (RUNDELL, 2008, p. 228).

---

[7]Not all CCELD's features were imitated: some were ignored, some rejected outright. Nevertheless, Sinclair, through CCELD, had a catalytic effect on lexicography, beginning with monolingual learners' dictionaries then spreading to bilingual dictionaries and monolingual dictionaries for adult or child native speakers, in Britain and internationally.

Baseados nos estudos de Kilgarriff, desde meados de 1980 com a possibilidade de versões de dicionários serem legíveis por máquina, surgiram programas experimentais para auxiliar no processo de desambiguação lexical de sentido, juntamente com o advento de corpora informatizados e o rápido crescimento do trabalho estatístico em Processamento de Língua Natural (PLN). Em uma de suas propostas (KILGARRIFF, 1997), atribuiu ao corpus a capacidade de desambiguar significados, demonstrando que as palavras possuem diversas acepções de acordo com o seu 'sentido atômico' (*core meaning*). Ele propõe, como alternativa, que as palavras sejam agrupadas de acordo com seus aspectos semânticos e com base na convergência estatística de seus variados usos que aparecem no corpus. Desse modo, a prática lexicográfica usou desses recursos para averiguar as inúmeras ocorrências de uma palavra no contexto, o que gerou mudanças na forma como lexicógrafos pensam sobre o significado.

Na concepção de Zampolli (1994), com a introdução de computadores, a palavra "lexicografia", que, até recentemente, tinha seu uso estabilizado no sentido de "a arte e a prática de fazer dicionário", tem se tornado um termo insuficiente para lidar com a realidade das atividades lexicográficas atuais. Krieger (2006) corrobora este pensamento, principalmente no que tange ao trabalho dos lexicógrafos. Segundo a autora:

> a constituição dos *corpora* informatizados agilizou o trabalho lexicográfico, que tem se beneficiado largamente do grande número de aplicativos que identificam rapidamente as palavras e seus percentuais de ocorrência. Em síntese, os princípios de seleção lexical, a escolha das fontes de coleta e o apoio logístico das tecnologias informatizadas identificam a nova fase da lexicografia que, na busca de cientificidade, mantém um diálogo incessante com a Linguística de Corpus. (KRIEGER, 2006, p. 145).

De maneira geral, os dicionários apresentam uma lista de potenciais significados enumeráveis para cada palavra, enquanto que, nos corpora, encontramos vestígios de tais significados de acordo com o contexto. Na maioria dos casos, os dicionários adotam o modelo aristotélico clássico de definições, que está baseado na distinção entre gêneros e diferenças. Contudo, há uma grande dificuldade ao escolher um termo que não seja nem muito geral, nem muito específico. Sendo assim, muitas vezes, os dicionários tendem a definir sua escolha pelo sinônimo e/ou antônimo, conforme menciona Fontenelle (2008, p.8). Com o advento da lexicografia baseada em corpus, essa área tem se expandido muito, principalmente devido às outras ferramentas que são utilizadas para averiguar a possibilidade das ocorrências nos mais diversos contextos e com seus respectivos empregos.

Os aspectos teóricos anteriormente mencionados estão intrinsecamente relacionados com a proposta da Lexicografia e das demais vertentes teóricas que a acompanham nesse processo, como, por exemplo, a Linguística, a Lexicologia, a Linguística de Corpus, entre outros. Tomados em sua totalidade, influenciam escolhas do lexicógrafo na concepção do produto dicionário. Atkins (2008, p.36) apresenta uma tipologia das decisões de design que podem estar envolvidas na concepção de um dicionário comercial – Figura 11. A cronologia do processo de tomada de decisão segue a linearidade proposta no gráfico, que progride da esquerda para a direita, ou seja, primeiro temos a fase pré-lexicográfica, depois a macroestrutura e, por último, a microestrutura.

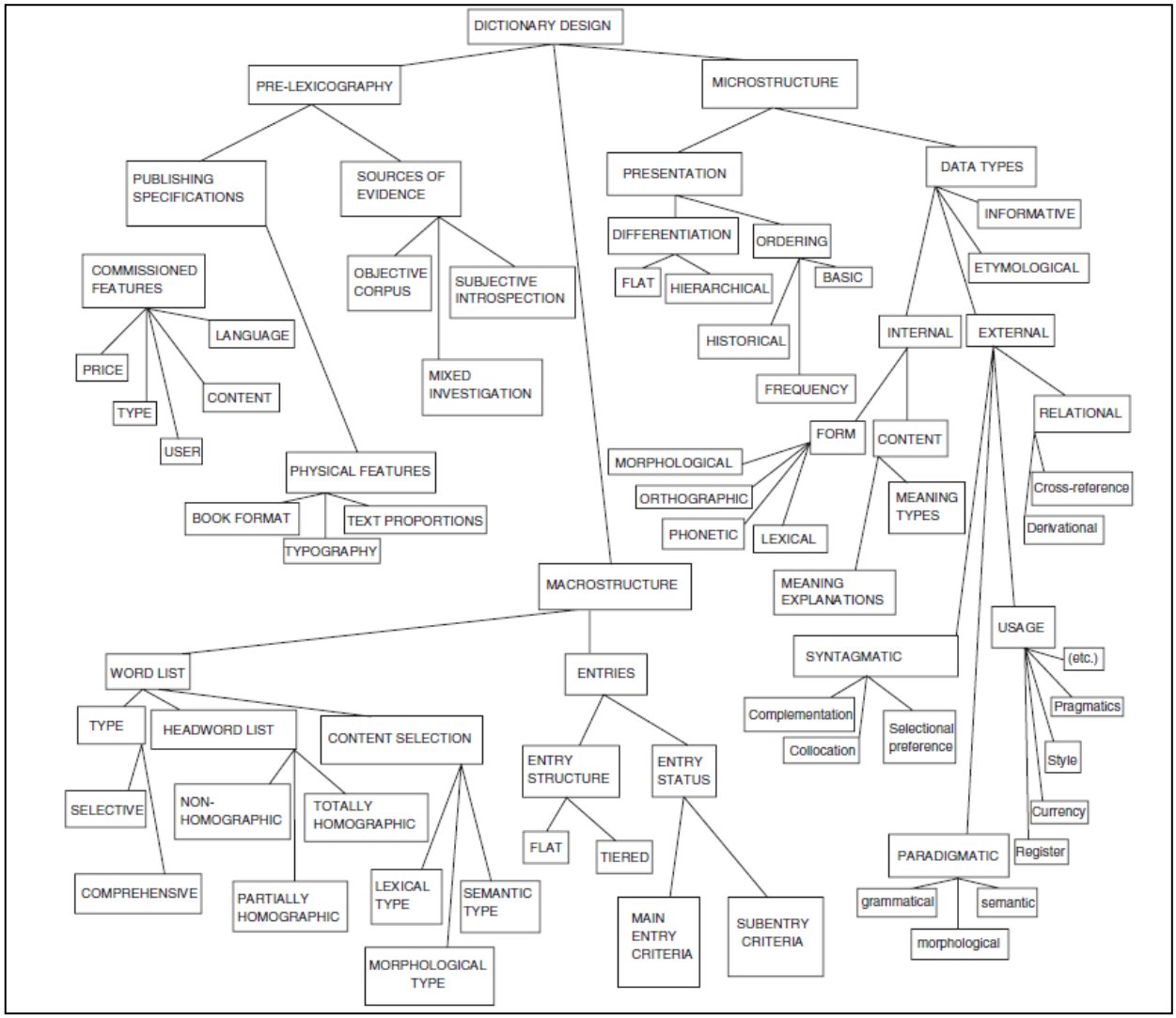


Figura 11: Gráfico que representa a tipologia das decisões de design de dicionário de comércio

Fonte: ATKINS, 2008, p.36

No primeiro momento, temos a chamada fase pré-lexicográfica, em que o lexicógrafo analisa a palavra para averiguar os fatos linguísticos relevantes e possíveis através de um processo de observação, com a finalidade de entender e ordenar o grau de conhecimento linguístico a ser representado pela obra. As fontes de evidências lexicográficas podem incluir um extenso corpus eletrônico, manuseado através de ferramentas sofisticadas de manipulação para ser anotado. Dessa forma, quanto maior o corpus, menor a probabilidade de o produto ser tendencioso pelo viés subjetivo, como ocorrem com alguns dicionários, que são escritos inteiramente por introspecção, sem o auxílio de citações ou de discussão com outras fontes. Como fator intermediário, temos a produção da maioria dos dicionários baseada na mescla das abordagens objetivas e subjetivas, que dependem de arquivos de citação e de informantes. Independentemente da dimensão do processo de análise, são fornecidos ao lexicógrafo os dados para a fase do processo, denominada por Atkins de síntese:

> O processo de síntese, a escrita real das entradas do dicionário, exige habilidades de interpretação, avaliação, seleção, compressão e comunicação, e uma certa quantidade de conhecimentos linguísticos. (ATKINS, 2008, p.34).[8]

A mesma análise pode gerar diversas sínteses, podendo ser cada uma delas um dicionário ou livro diferente. No decorrer da etapa de síntese, o compilador extrai a partir dos dados coletados os fatos ordenados que são relevantes para o dicionário que está sendo escrito. Ainda neste processo, pode haver uma grande contribuição do linguista teórico quando este introduz as competências linguísticas lexicográficas e a sensibilização para as línguas, uma vez que o processo de síntese requer do lexicógrafo alguns procedimentos especiais, como, por exemplo, verificar para quais tipos de tarefas o dicionário será usado. Assim, ele terá uma ideia mais clara das expectativas, das necessidades e dos possíveis problemas que o usuário pode encontrar; assim como averiguar as restrições comerciais, tipográficas (fonte, tamanho) que envolvem o espaço (quantidade de páginas, formato), o tempo a ser gasto na produção, os custos, o tamanho da equipe envolvida, o processo de compilação, se será linguagem geral ou especializada, se será monolíngue, bilíngue ou multilíngue, e o impacto sobre o processo editorial final, que podem acarretar no sucesso ou fracasso do dicionário.

---

[8] The synthesis process, the actual writing of the dictionary entries, calls for skills of interpretation, evaluation, selection, compression, and communication, and a certain amount of linguistic knowledge.

As principais tomadas de decisão na fase pré-lexicográfica são relativas às especificações e publicação, sendo tais decisões tomadas não pelos editores do dicionário e sim pelos editores de mercado. Desse modo: “os editores de um novo dicionário vão avaliar cuidadosamente a possibilidade de mercado combinando o produto aos usuários previstos para garantir um bom ajuste.” [9] (ATKINS, 2008, p.37). Em suma,

> Enquanto a lexicografia teórica tem uma importante contribuição a dar para a pré-lexicografia no âmbito da coleta e pré-processamento de provas lexicográficas, nem os linguistas nem os lexicógrafos têm qualquer palavra a dizer nas decisões que afetam a lexicografia mais radicalmente, ou seja, as decisões de negócios feitas por editores. (ATKINS, 2008, p. 39)[10]

Outro aspecto interligado às especificações e publicações são as características físicas que geralmente são parcialmente definidas na fase de pré-planejamento, englobando decisões a respeito do formato do livro (tamanho, tipo de cobertura, etc), tipografia (tipos de fonte, tamanhos, etc) e as proporções de texto (quantidade de páginas disponíveis para o dicionário).

O último item nesta fase é relativo à especificação do conteúdo, que segundo ATKINS (2008, p. 38), não é explícito nesta fase inicial, sendo apenas o tipo de conteúdo definido neste momento. Em suma, a fase pré-lexicográfica envolve os critérios de espaço (quantidade e tamanho das entradas) e tempo (tempo disponível para compilação e análise do texto).

Na segunda etapa, seguindo a ordem apresentada no gráfico, temos a macroestrutura, que é considerada a base do dicionário, ou seja, tudo que molda a informação linguística. Estas decisões se dividem em dois grupos: as que se referem à lista de palavras e as relativas às entradas lexicais. No que concerne ao primeiro grupo, os editores irão decidir se a lista de palavras será mais abrangente ou seletiva. Em seguida, os editores deverão então resolver como estruturar o vocabulário a ser tratado no texto do dicionário, escolhendo qual tipo de lista será usado como critério, podendo ser não-homográfico (uma única forma ortográfica e todas as unidades lexicais com a mesma forma consideradas como constituintes de uma única entrada polissêmica);

---

[9]The editors of a new dictionary will carefully assess the market slot matching the product to the intended users to ensure a good fit.

[10]While theoretical lexicography has an important contribution to make to the pre-lexicography in connection with the gathering and pre-processing of lexicographical evidence, neither linguists nor lexicographers have any real say in the decisions that affect the lexicography most radically, namely, the business decisions made by publishers.

assim como também pode ser totalmente homográfico, onde cada verbete é uma única unidade de forma e conteúdo e uma única forma ortográfica podendo ser compartilhada por muitas palavras chaves, nenhum dos quais é polissêmico; e, por último, parcialmente homográfico, que é utilizado nos dicionários comerciais, inclusive os dois exemplares discutidos nesta seção, onde um único formulário pode ser compartilhado por várias entradas, cada um dos quais podendo ser ela própria polissêmica (ATKINS, 2008, p. 39, 40).

Outra configuração importante na macroestrutura é o estabelecimento de critérios para constituir a seleção do conteúdo real da lista. No primeiro conjunto, temos os critérios relacionados ao tipo lexical (palavras inteiras, sem itens polilexêmicos); o segundo conjunto de critérios relaciona-se com tipo morfológico, especificamente incluindo ou excluindo derivados ou formas flexionadas, afixos, etc., e o terceiro conjunto de critérios refere-se ao tipo de semântica, a seleção de itens de inclusão ou exclusão, como nomes próprios, palavras tabu e assim por diante.

O segundo grupo de decisões que afeta a macroestrutura do dicionário são as entradas reais, podendo ser planas – ou seja, caso em que todos os sentidos irão apresentar um status igual, no qual serão sequenciados numericamente em 1,2,3,4..., de acordo com a proximidade do significado em que cada item tem sua própria entrada lexical – ou em camadas – no qual o significado pode ser agrupado de maneira hierárquica, o que permite o agrupamento de sentidos de uma forma mais intuitiva como, por exemplo,1,1a,1b, 2. Neste caso, temos uma subdivisão dentro de uma mesma entrada, ou dentro de outras sub-entradas. As sub-entradas podem conter outros tipos de informações que podem ser sistematicamente incluídas ou excluídas, de acordo com critérios gramaticais, morfológicos ou ortográficos.

Na terceira etapa, temos as decisões sobre a microestrutura, que está mais voltada para os aspectos linguísticos do que propriamente os lexicográficos. Os editores deverão escolher o tipo de apresentação das unidades lexicais, que podem ser definidas pela diferenciação de sentido, que pode ser plana ou em camadas, e a ordem como estes serão definidos na entrada do dicionário; a ordenação dos sentidos também deve ser feita pelo editor na fase de concepção e será definido se seguirá a evolução histórica dos sentidos, ou a sua frequência relativa percebida, ou se deverá ser executada a partir do sentido intuitivo que se acredita ser fundamental.

Os designers do dicionário são responsáveis pelos dados lexicalmente relevantes, que são divididos em quatro tipos diferentes de informações: (i) interna, sobre a palavra

ou o próprio sentido da palavra, ou seja, diz respeito a sua forma ou sentido; (ii) externa, sobre as suas relações com outras palavras ou sentidos da palavra; (iii) etimológica, de certa forma um subconjunto diacrônico das externas; e (iv) informativa, uma coleção heterogênea de comentários úteis que o compilador pode fazer a fim de esclarecer uma entrada particular (ATKINS, 2008, p. 43).

Nas informações internas encontramos dados referentes à base morfológica, que sempre é dada, podendo apresentar também as formas flexionadas; outro aspecto presente é o ortográfico, condizente com a grafia padrão e as das variantes e uma possível silabação. Pode haver também informações fonéticas e, por último, lexicais, que se caracterizam por indicar se trata de um único item lexical, se é uma palavra com hífen entre outros. Quanto ao conteúdo semântico, este pode ser analisado sob o viés que abrange os vários tipos de significado, podendo ser incluído na entrada com diversos tipos, como denotação, conotação, referência, significado literal ou figurado e assim por diante. O segundo aspecto abarcado pelo conteúdo semântico é a explanação dos significados que constitui na entrada não apenas a definição, mas também comentários, exemplos de usos, notas e referências cruzadas chamando atenção do usuário para as entradas relacionadas.

No que tange às informações externas, Atkins (2008, p.46) assinala que se trata dos fatos que um dicionário pode apresentar relacionando a entrada com outras palavras do dicionário. A autora apresenta quatro vertentes: na parte concernente à informação relacional, temos vários tipos de referências cruzadas para as outras entradas do dicionário, que consistem no conjunto de fatos demarcados pela derivação morfológica. Na relação paradigmática, podemos encontrar as informações gramaticais, semânticas e morfológicas. A parte considerada mais complexa é a que se refere às informações de uso, ou seja, aos fatos que serão rotulados no dicionário, como o tipo de registro que será adotado (como, por exemplo, formal, informal), estilo (técnico, poético), campo (música, arquitetura), pragmática (expressa surpresa), status (dialetos, gírias), entre outros.

De acordo com Atkins (2008, p. 47) os maiores problemas do editor com os fatos externos se encontram nas relações sintagmáticas, na fase de concepção de um projeto de dicionário. O que pode ser considerado um grande benefício, também pode trazer alguns problemas para as restrições seletivas, devido ao uso de corpora eletrônico como prova lexical. A riqueza de informações e citações disponíveis podem trazer mais perguntas do que respostas. As evidências em corpus das *collocations* podem designar

a co-ocorrência significativa de duas ou mais palavras (ou porque tal agrupamento acontece com mais frequência do que a norma estatística, ou porque ele representa um item lexical semanticamente opaco). Para Atkins,

> A linguística teórica abunda em escritos sobre a fraseologia; não há escassez de comentários interessantes e perspicazes sobre este aspecto da linguagem. Mas o desafio para o teórico ainda está em aberto: como pode um editor classificar as colocações da língua, e, tendo feito isso, elaborar critérios para a atribuição de cada item polilexêmico encontrado na sua classe apropriada? (ATKINS, 2008, p. 47).[11]

Um exemplo que Atkins (2008, p. 48) aborda é a dificuldade que os lexicógrafos encontram para decidir quais usos verbais com preposição devem ser tratados como verbos frasais e, portanto, receber um tratamento especial como unidades lexicais. Segundo a autora este problema ainda continua em aberto, ou seja, sem solução.

Na tentativa de explicar a importância da inter-relação entre lexicógrafos e áreas afins, Atkins separa em três grupos os linguistas teóricos que poderiam ou não se considerar lexicógrafos teóricos, sendo eles: (i) aquele que trabalha mais essencialmente com a parte teórica, ou seja, uma parte mais abstrata, distanciando-se da concepção do que seja lexicografia prática; (ii) aquele que discute aspectos da semântica lexical (a intersecção existente entre semântica e sintaxe), fator relevante e de grande influência na lexicografia prática, e por último (iii), o que trata de assuntos especificamente lexicográficos, de maneira que nem sempre são aplicáveis à lexicografia, como os sociolinguistas, psicolinguistas, especialistas em estilística e análise do texto, que colaboram com uma abordagem estruturada para a entrada lexical e o seu conteúdo.

No capítulo 4, onde apresentaremos a Diciopédia, uma das ferramentas do m.knob e nosso objeto de estudo, retomaremos os critérios que foram discutidos neste capítulo, para descrever as principais decisões lexicográficas concernentes à elaboração da Diciopédia e de sua composição.

Às peculiaridades que foram listadas acima, precisamos acrescer outros detalhes e particularidades no que se refere aos dicionários e glossários de especialidade, que procuram abarcar domínios e temas pontuais. Assim, na próxima seção, abordaremos a Terminologia.

---

[11]Theoretical linguistics abounds in writings on phraseology; there is no dearth of interesting and perspicacious commentaries on this aspect of language. But the challenge to the theoretician is still open: how may an editor classify the collocations of the language, and, having done so, devise criteria for assigning each multiword item encountered to its appropriate class?

1.2 Terminologia

A terminologia de uma área de conhecimentos é definida como um conjunto de termos de um sistema global do léxico, perfilado no interior de uma língua, que trata de um âmbito especializado, no qual encontramos vocabulário específico que pode ser subdividido em grupos temáticos ou em campos semânticos e pragmáticos (HOFFMANN, 2004, p. 85).

Enquanto nome de uma ciência, Terminologia refere-se a uma teoria interdisciplinar, uma vez que as características das unidades terminológicas estão intrinsecamente relacionadas com a linguagem (Linguística), com elementos da cognição (Ciência Cognitiva), com veículos comunicativos (Comunicação), com comunicação especializada (Documentação) e também com a Informática (Terminografia).

De acordo com Cabré (1999, p. 27), a origem da Terminologia data da década de 30, com os trabalhos de Eugen Wuster em Viena, que teve como preocupação inicial os aspectos normativos e metodológicos, o que justificou seus métodos de trabalho e os princípios que deveriam presidir os trabalhos sobre os termos[12], dando origem à Teoria Geral da Terminologia (TGT). A segunda etapa foi denominada de estruturação (1960-1975), que foi marcada pelo desenvolvimento da macroinformática e das técnicas documentais. Nesse momento, aparecem os primeiros bancos de dados e a organização internacional de Terminologia. Na etapa seguinte, o período foi nomeado de eclosão (1975-1985), na qual tivemos o processo de modernização dos repositórios terminológicos de uma língua, principalmente devido à expansão da microinformática. No último período, a partir de 1985, conhecido como ampliação, o avanço da informática possibilitou o surgimento e o aprimoramento de novos recursos de trabalho, sendo estes mais eficientes para os terminólogos. Foi um período de consolidação de um modelo de terminologia ligado à planificação da língua.

---

[12]Segundo Cabré (1999,p.87), “Los términos suelen tener en teoría un significado, solamente uno, e inequívoco, puesto que aquellos se describen en el interior de un campo de especialidad concreto, y no en abstracto, sin especificar los ámbitos de uso, como ocurre con el repertorio léxico general”.

Para Cabré (1999, p. 33), a Terminologia se preocupa prioritariamente com a relação existente entre o objeto da realidade e o conceito que o representa, ou seja, está estreitamente vinculada aos campos de especialidade dos quais se ocupa. Sendo assim, a autora considera que:

> A terminologia não é um objeto que se justifique por si mesmo, nem o trabalho terminológico pode ser considerado como uma mera recompilação de uma série de conceitos com suas denominações, sem outra finalidade. A terminologia serve a ciência, a técnica e a comunicação e deve ser consequente com essa função. Especialistas de cada domínio temático e especialistas em terminologia geral e aplicada trabalham em colaboração para levar a cabo as tarefas de ordenação e normalização dos conceitos e termos de cada disciplina.[13]

Neste sentido, tomando como base essa tridimensionalidade das faces constitutivas da Terminologia (ciência, técnica e comunicação), Cabré (1999, p. 37), separa em duas perspectivas a funcionalidade de tal teoria: (i) para aqueles que a utilizam para comunicação direta ou através de intermediários; (ii) e aqueles que trabalham com ela, com finalidade de produzirem glossários, ou de auxiliarem na informação através de alguma ferramenta.

Na concepção de Cabré (1999, p. 82), a Terminologia denomina um objeto conceptual específico, descrevendo-o exaustivamente através de uma definição, expressando suas relações com conceitos distintos. Além disso, ordena sistematicamente suas entradas, ou seja, em ordem alfabética, permitindo uma apresentação mais adequada dos dicionários multilíngues, já que existe uma teoria única para a noção especializada, diferindo-se apenas na forma de denominação. Dessa forma, a autora esclarece que:

> A terminologia concebe os termos em um sentido internacional e, por isso, tende a privilegiar os modos de formação que, nas linguagens de especialidade, aproximam as línguas históricas entre si. Tal fenômeno conduz à adoção de critérios internacionais de formação de termos e a orientações de trabalho cuja validade embasa o âmbito de cada língua particular.[14]

---

[13] La terminología no es un objeto que se justifique por sí mismo, ni el trabajo terminológico puede quedarse en una mera recopilación de una serie de conceptos con sus denominaciones, sin otra finalidad. La terminología sirve a la ciencia, a la técnica y a la comunicación, y debe ser consecuente con esa función. Especialistas en cada dominio temático y especialistas en terminología general y aplicada trabajan en colaboración para llevar a cabo las tareas de ordenación y normalización de los conceptos y términos de cada disciplina.

[14] La terminología concibe los términos en un sentido internacional, y por ello tiende a privilegiar los modos de formación que, en los lenguajes de especialidad, acercan las lenguas históricas entre sí. Dicho fenómeno conduce a la adopción de criterios internacionales de formación de términos, y a guías de trabajo cuya validez rebasa el ámbito de cada lengua particular.

Na concepção de Hoffmann (2004, p. 86), o trabalho terminológico apresenta como principais características: (i) delimitar a correspondência entre um conceito e um signo linguístico; (ii) estabelecer relações hierárquicas e conceituais, como de gênero-espécie, causa, entre outras; (iii) determinar, dividir e organizar as características essenciais, que são de grande relevância na definição dos conceitos; (iv) otimizar a comunicação especializada de modo a eliminar mal entendidos, assegurando ainda mais a comunicação entre os especialistas, e (v) auxiliar no ensino instrumental de línguas estrangeiras através de um componente essencial do vocabulário especializado, com ênfase para o papel do léxico.

Segundo Krieger e Finatto (2004, p. 17), o objeto primordial da Terminologia é o termo técnico-científico, que apresenta dois enfoques distintos: o desenvolvimento teórico juntamente com as análises descritivas e as aplicações terminológicas, que compreendem glossários, dicionários especializados, banco de dados terminológicos e sistema de reconhecimento automático de terminologias. Essa dupla face configura a identidade da disciplina Terminologia. Assim, os termos transmitem conteúdos específicos de cada área, apresentando dessa forma duas funções primordiais: a de representação e a de transmissão de conhecimento especializado.

Essas categorias apresentadas estão mais interligadas à Terminologia Convencional, que tem como alvo a compreensão e aquisição de conhecimento especializado, assumindo a forma de árvores conceituais (FABER ET AL., 2005). Outra corrente teórica é a Terminologia Sociocognitiva (TEMMERMAN 1997, 2000), que apresenta um vínculo estreito com os fatos da língua e os conceitos, procurando mostrar as relações entre os processos de categorização e a linguagem, acolhendo os fenômenos de realização de variação, de sinonímia e polissemia na linguagem especializada, que até então eram rechaçados pelos estudos clássicos de Terminologia; assim como consideram que os conceitos e os termos evoluem com o tempo, priorizando o uso de ontologias no processo de criação de recursos terminológicos. Nessa concepção, os termos são considerados unidades de compreensão e representação, configurando-se como modelos cognitivos e culturais, designando categorias da estrutura prototípica de maneira claramente delimitável.

Há também, nesta área, o enfoque teórico-prático denominado Terminologia baseada em Frames, doravante TbF, (FABER ET AL., 2005, 2006, 2007), que compartilha muitos aspectos da Terminologia Convencional e da Terminologia Sociocognitiva, tais como os que foram mencionados anteriormente. Um dos principais

objetos de estudo da TbF são as unidades lexicais e terminológicas, que são determinadas de acordo com o contexto e estão intrinsicamente relacionadas a um domínio especializado. Tais unidades são consideradas como configurações de eventos complexos e estão interligadas através dos conceitos do domínio. Além disso, estão ancoradas em dados empíricos fornecidos pela Linguística de Corpus.

Segundo Faber e colaboradores (2005), as áreas especializadas devem ser organizadas em frames[15], de modo que estes forneçam tanto bases conceituais para a localização de conceitos hierárquicos dentro de um domínio especializado, quanto de um modelo de definição.

Embora a Terminologia tenha progredido ao longo dos anos, alguns aspectos ainda são considerados desafiadores nesta área, conforme elucidam López-Rodriguez e colaboradores (2010):

> Um dos desafios atuais da Terminologia é que, diante de uma visão tradicional prescritiva da terminologia, a prática terminológica e o uso real dos termos têm mostrado que é necessário um enfoque descritivo que estude as unidades de significação especializada desde uma perspectiva social, linguística e cognitiva, que leve em conta o dinamismo da língua. Em segundo lugar, os novos suportes de multimídia da informação e da interdisciplinaridade da Ciência e da Técnica têm gerado a necessidade de um formato e um modelado adequados para a descrição multidimensional e multimodal dos termos (LÓPEZ-RODRIGUEZ ET AL., 2010, p. 54).[16]

Segundo os autores citados anteriormente, para essa descrição multidimensional e multimodal, os recursos lexicais e terminológicos precisam incluir informações linguísticas (textual e lexical), assim como imagens, áudio, vídeos. Outro aspecto muito relevante é a forma como os conceitos são representados, uma vez que isso afeta a configuração da informação dentro das entradas terminológicas individuais e os conteúdos que estão relacionados no banco de dados, especialmente, ao que diz respeito à descrição ou definição conceitual (FABER ET AL., 2006, p. 191).

---

[15]Segundo Fillmore (1982, p. 111), um frame também pode ser definido como “qualquer sistema de conceitos relacionados de tal forma que, para entender um deles, é necessário compreender toda a estrutura na qual ele se encaixa”. A seção 2 se dedicará à Semântica de Frames.

[16]Uno de los desafíos actuales de la Terminología es que, frente a una tradicional visión prescriptiva de la terminología, la práctica terminológica y el uso real de los términos han mostrado que es necesario un enfoque descriptivo que estudie las unidades de significación especializadas desde una perspectiva social, lingüística y cognitiva, que tenga en cuenta el dinamismo de la lengua. En segundo lugar, los nuevos soportes multimedia de la información y la interdisciplinariedad de la Ciencia y la Técnica han generado la necesidad de un formato y un modelado adecuados para la descripción multidimensional y multimodal de los términos.

Dessa forma, a construção de um modelo de evento para um domínio de conhecimento especializado ajuda ao usuário a processar o conteúdo conceitual com mais facilidade, pois, baseadas na gestão da terminologia, as informações essenciais serão configuradas através de redes, que, por sua vez, são divididas em domínios e estes, em frames, que poderão passar por vários níveis de especificidade usando herança hierárquica. Trata-se de um aspecto considerado flexível, uma vez que os eventos são dinâmicos e entidades do mundo real podem desempenhar papéis diferentes neles.

Na apresentação do projeto EcoLexicon e do aplicativo m.knob, retornaremos à temática da TbF. Por ora, entretanto, para compreendermos melhor a face aplicada da Terminologia, correlacionando-a com a Lexicografia, dedicaremos a próxima seção à Terminografia.

1.2.1 Terminografia

A Terminografia é a vertente prática da Terminologia, que se encarrega da produção de dicionários especializados e também de estudos sobre termos. Segundo Cabré (1999, p.263), a atividade terminográfica integra operações de recompilação, sistematização e apresentação dos termos de uma determinada área do saber humano. Além disso, a elaboração de dicionários leva diretamente à normalização de termos próprios de um determinado domínio especializado.

Existem alguns princípios na atividade terminográfica que precisam ser estabelecidos com o propósito de instituir uma prática rigorosa, sistemática e controlada pelas bases explícitas de uma teoria sobre os termos, conforme elenca Cabré (1999, p. 266), sendo eles: (i) a consideração dos termos como unidades indissociáveis de uma vertente dupla, que considera forma e conteúdo; (ii) a tendência de manter uma relação unívoca entre a forma e o conteúdo de um termo; (iii) a localização necessária de um termo em um campo conceitual determinado, sem referência ao qual não teria um valor específico; (iv) a relação inevitável de cada conceito com os demais conceitos que constituem um campo específico; (v) o conjunto de possibilidades estruturais que possuem as denominações.[17]

---

[17] Estas recomendações e princípios foram estipulados pelo Comité Técnico 37 (CT37) da ISSO (CABRÉ, 1999, p. 267).

Em virtude das analogias, semelhança formal e paralelismo nos produtos dicionarísticos, a Lexicografia e a Terminografia podem ser confundidas. Por isso, Cabré (1999, p. 92) esclarece algumas de suas principais diferenças, dentre elas, temos que a confecção de um dicionário de léxico geral parte de uma lista de palavras que constitui o inventário das entradas de um dicionário. O lexicógrafo continua a descrevê-las semanticamente por meio da definição, sendo considerado de tipo semasiológico, ou seja, se passa da forma para o significado. Por outro lado, a Terminografia trabalha de maneira inversa, já que o terminológo, primeiramente, estabelece a lista de conceitos que formam parte de uma estrutura nocional. Estes, por sua vez, estarão inter-relacionados lógica e ontologicamente, tendo em vista que o conjunto compõe o sistema conceitual de uma atividade especializada. O terminógrafo atribui a cada espaço conceitual uma determinada denominação, que irá se referir a uma dada estrutura nocional correspondente à forma que se utiliza efetivamente. No caso de confluência entre várias denominações para um mesmo conceito, se seleciona uma delas, descartando as demais, ou se elege uma delas, configurando-a como prioritária, aceitando as outras possibilidades. Este processo é denominado de onomasiológico, visto que há um procedimento básico de identificação dos termos ou dos aspectos conceituais, no qual buscam-se os termos, que serão entendidos como etiquetas denominativas de tal conteúdo.

Krieger e Finatto (2004, p.54) elaboraram um quadro comparativo, que elucida de maneira mais objetiva as principais diferenças entre essas duas disciplinas, conforme Figura 12.

	Lexicografia	Terminologia
Vertentes	Prática e teórica	Prática e teórica
Origem	Tradição	Contemporaneidade
Objeto	Léxico geral/ palavra	Léxico temático/ termos
Produto	Dicionários de língua mono, bi e multilíngües	Léxico, glossários, dicionários terminológicos mono, bi e multilíngues, banco de dados terminológicos.
Natureza	Linguístico-descritiva	Cognitiva-normalizadora

Objetivos e funções	Repertoriar o léxico geral; oferecer informações etimológicas, gramaticais, sociolinguísticas; oferecer informações semânticas gerais e especializadas de todas as unidades lexicais de um idioma (polissemia); oferecer padrões de usos linguísticos; legitimar o léxico de uma língua.	Repertoriar o léxico temático; oferecer informações terminológicas e conceituais de uma área de conhecimento especializado; delimitar conceitos de um sistema cognitivo específico (homonímia); estabelecer padrões de designação e conceituação em áreas de especialidade (normatização).
Usuário	Difuso	Específico
Fontes	Textos em geral	Textos de especialidade
Método	Semasiológico	Onomasiológico
Entradas		
Critério de seleção	Frequência	Pertinência do termo à área de conhecimento/ frequência em menor escala
Tipologia	Verbal: palavras gramaticais e lexicais	Verbal: termos simples, compostos, siglas e acrônimos. Não-verbal: símbolos e fórmulas.
Tratamento	Lematização, forma canônica.	Manutenção da forma plena e recorrente.

Figura 12:Comparativo entre Lexicografia e Terminologia

Fonte: KRIEGER e FINATTO, 2004,p.54

Os principais materiais constituintes do trabalho terminográfico usados como suporte, referência e armazenamento de informações são: materiais de consulta, que auxiliam nas informações sobre os distintos aspectos terminológicos ou de uma área especializada; materiais específicos de trabalho (básico); materiais de suporte ao trabalho que completam e facilitam o desenvolvimento da elaboração de um glossário (CABRÉ, 1999, p. 270).

Outro aparato de extrema importância no trabalho terminográfico é o uso de corpora textuais, pois, quando o contexto de um item lexical é explicitado, temos o seu significado interligado ao seu uso. Nesse sentido, L'Homme e colaboradores (1999, p.32-33) apontam que o contexto ajuda a recuperar informações sobre termos relacionados, características de definição, sinônimos, bem como as relações explícitas entre conceitos semanticamente relacionados. Faber e outros (2005), sob esse enfoque, acrescentam que:

> Além disso, com base no contexto, é possível verificar a informação conceptual e gramatical extraída de livros de referência, especificar o significado de abreviaturas e também verificar a informação enciclopédica para incluir na definição.[18]

Na concepção de Cabré (1999, p. 278), a seleção das fontes é efetuada para concretizar determinado trabalho, pois, a partir desse corpus, serão extraídas denominações que integrarão a primeira lista de trabalho de recompilação de uma disciplina. Dessa forma, podemos entender a importância dos corpora como fonte de seleção dos termos, por isso, segundo Cabré (1999, p. 298), o corpus deve cumprir uma série de condições para garantir a confiabilidade dos resultados, tais como: ser pertinente, isto é, representativo do campo pesquisado; ser completo, incluindo, assim, todos os aspectos que devem estar relacionados com o tema investigado; ser atual, para que a lista de termos extraída reflita a realidade linguística do âmbito em questão; e, por último, ser original, ou seja, estar expresso na língua com que se pretende trabalhar.

O trabalho terminológico realizado com base em corpora automatizados tem possibilitado muitas vantagens, que, algumas décadas atrás, não eram possíveis, como manusear maior quantidade de textos e dados, conferindo maior flexibilidade e uma projeção múltipla destinada a diferentes usuários, além de proporcionar o armazenamento em grandes bancos de dados.

Os princípios delineados acima aplicam-se à constituição de dicionários de domínio específico, aos quais será dedicada a próxima seção.

1.2.2 Constituição de Dicionários de Domínio Específico

Na elaboração de um dicionário terminológico, o processo é longo e requer uma atenção especial com a nomenclatura, ou seja, com a lista de verbetes que irão compor o dicionário, tendo em vista que se trata de uma situação comunicativa particularizada, com especificidades distintas das de um dicionário geral ou de língua. Nesse sentido, para que seja um produto com êxito, que atenda às demandas do usuário, é necessário adotar metodologias de pesquisa coerentes e que correspondam ao uso real do consulente.

---

[18]Furthermore, on the basis of context it is possible to verify the conceptual and gramatical information extracted from reference books, specify the meaning of abbreviations, and also to verify encyclopaedic information to include in the definition.

Na etapa de planejamento, Krieger e Finatto (2004, p. 128) destacam as seguintes situações que devem ser observadas com antecedência, em consonância com o que pontua Castillo (1997): verificar as características da definição que constarão no verbete, assim como as delimitações de termos, priorizando a predominância de sintagmas nas terminologias; averiguar os modos de apresentação dos equivalentes de tradução em outras línguas; prever espaços para a indicação de neologismos, estrangeirismos, polissemia e variação terminológica; disponibilizar ferramentas informatizadas para auxiliar o processo de automação e agilização das tarefas; seguir os padrões de obras semelhantes à temática a qual se pretende reproduzir. É importante ressaltar a importância do envolvimento de todo o grupo de trabalho, uma vez que há diversas tarefas a serem desenvolvidas.

Dentre as várias etapas que compõem a constituição de dicionário específico, Krieger e Finatto (2004, p. 129) assinalam que o reconhecimento terminológico é o primeiro passo concreto. Para isso, é relevante observar se as unidades linguísticas registradas num acervo mantêm relações tanto sintagmáticas (colocações) como paradigmáticas (sinonímia, antonímia, etc.) com outras unidades, uma vez que “relações de sentido e pertinências devem ser levadas em conta no momento do reconhecimento que tem por objetivo a composição de um glossário” (KRIEGER E FINATTO, 2004, p. 129).

Outro aspecto importante é averiguar se as unidades que se registram como termos são realmente representativas para a área do saber para a qual se organiza o dicionário. Além disso, a adoção de métodos e a organização devem seguir os critérios e os padrões dos fundamentos da Terminografia, estabelecidos pela ISO (*International Standardization Organization* – instituição dedicada às normalizações técnicas em geral) com o intuito de facilitar a transferência e o acesso por equipes diferentes.

Para saber se um determinado conjunto de expressões está vinculado a conceitos de uma dada área de especialidade, observa-se o registro delas em corpora terminológicos, como ressaltam Krieger e Finatto (2004, p. 129):

> Essas expressões são identificadas e coletadas em corpora terminológicos, isto é, são extraídas de uma base textual representativa, fixada por critérios previamente determinados (Picht, 2001). Portanto, o reconhecimento terminológico, frisamos, está intrinsecamente relacionado ao reconhecimento de textos técnicos ou científicos e à identificação de tipos textuais, sejam eles mais ou menos “especializados” ou mais ou menos terminologicamente densos. Vistas em textos, as expressões de valor terminológico podem ser puramente verbais (palavras ou grupos de palavras), não verbais (símbolos) ou, ainda, de natureza mista (símbolos e palavras). (KRIEGER e FINATTO, 2004, p. 129).

As autoras ressaltam que em dicionários e glossários o foco tende a ser em expressões nominais, denominadas de termos técnico-científicos, já que sua função é nomear o objeto de uma realidade, sendo mais numerosos do que os formados por uma só palavra. A classe gramatical que mais predomina no conjunto de entradas é o substantivo. Na concepção de Hoffmann (2004, p. 84), no vocabulário especializado predominam substantivos e adjetivos em relação aos verbos e outras classes de palavras, pois é preciso designar a multiplicidade de objetos e manifestações, objetos da atividade especializada. Por outro lado, vale mencionar que adjetivos, verbos, sintagmas terminológicos e fraseológicos também são elementos que integram uma linguagem técnico-científica. Segundo apontam Faber e colaboradores (2005), grande parte do nosso conhecimento é composto por muitos eventos e estados, dessa forma, temos uma grande representação linguística por meio dos verbos, desempenhando, assim, um papel crucial na linguagem.

Na escolha dos termos, privilegiam-se os que são mais empregados na atualidade. Os verbetes podem ser organizados em ordem alfabética, o que é mais comum, ou podem ser listados de acordo com uma ordem temática ou sub-temática, baseada na estruturação de um domínio conceitual específico concebido por um especialista da área. Já segundo Maciel (1996), os termos podem ser repertoriados também por pertinência pragmática, ou seja, termos que são originados de áreas afins e que contribuem para entender melhor os conceitos cobertos pela terminologia específica também podem ser elencados. Nesse caso, o contexto comunicativo é muito representativo, já que nele temos a manifestação da pertinência pragmática.

Dados os pressupostos teóricos para a seleção terminológica, Drouin (2003) apresenta uma metodologia de extração de termos a partir da comparação entre um corpus técnico e um não especializado. Seu escopo principal é diminuir a quantidade de “ruídos”, ou seja, de termos que não são correspondentes à área especializada em análise. Assim, espera-se que, a partir desse processo de extração de termos candidatos, o teste estatístico selecione itens lexicais altamente específicos para o corpus técnico. A ferramenta desenvolvida por Drouin (2003) é denominada TermoStat, software destinado à extração de termos complexos, e também de termos simples, que muitas vezes são ignorados pelos sistemas automatizados.

Em primeiro lugar, o autor utiliza uma técnica que visa a extrair os termos especializados de um corpus global (CG) que não é de linguagem técnica, sendo construído a partir de um corpus de referência (CR). As frequências desses termos são

comparadas estatisticamente com as extraídas de um corpus técnico, denominado de corpus de análise, doravante CA. Segundo Drouin (2003), *"tradicionalmente, o léxico específico foi identificado com base em uma comparação da frequência de unidades num subcorpus com a sua frequência em todo o corpus".*[19] O resultado obtido a partir da comparação das frequências pode possibilitar três situações, de acordo com Drouin (2003): (i) a frequência observada pode ser equivalente à frequência teórica; (ii) ela pode ser significantemente superior, (iii) ou inferior.

Contrastando dois corpora com conteúdos distintos, podemos comparar o comportamento das unidades lexicais em diferentes tipos de corpora dentro do corpus global, construído de maneira dinâmica, o qual possibilita identificar os itens lexicais que são específicos para o corpus de análise. Serão considerados apenas os itens que aparecem no corpus de análise com maior frequência do que aparecem no corpus global. Em suma, a intenção do autor é

> (...) usar a dicotomia observada entre o CR e o CA no nível do tópico para adquirir unidades lexicais específicas de domínio. Ao fazer isso, estamos usando informações sobre o comportamento de palavras em um corpus não técnico para ter acesso aos termos em um corpus técnico.[20]

Para quantificar a especificidade dos itens lexicais, dois critérios são adotados: (i) o valor do teste, que é considerado uma visão padronizada da frequência das unidades lexicais, e (ii) a probabilidade de observar um item com frequência igual ou superior ao apresentado no corpus de análise.

É importante mencionar que existe uma restrição para extrair os termos: é permitido que os algoritmos recuperem apenas substantivos e adjetivos (os chamados de pivôs lexicais específicos do corpus).

Apesar de a metodologia apresentada por Drouin (2003) ser interessante, existe uma fronteira muito tênue entre linguagem comum e linguagem especializada. Krieger e Finatto (2004, p.132) afirmam que "muitos termos de áreas específicas figuram em DLGs (dicionário de língua geral)". Por outro lado, a inserção de palavras não "técnicas" no repertório de uma área específica tem sido ainda muito discutida por especialistas, lexicólogos e terminólogos.

---

[19]Traditionally, the specific lexicon was identified based on a comparison of the frequency of units in a subcorpus compared to their frequency in the whole corpus.

[20](…) to use the dichotomy observed between the RC and the AC at the topic level to acquire domain-specific lexical units. In doing so, we are leveraging information about the behavior of words in a non-technical corpus to gain access to terms in a technical corpus.

Há também discussões a respeito da exclusão de palavras já definidas no DLG, devido a uma “delimitação” entre Lexicografia e Terminografia. Esses questionamentos refletem as dificuldades relacionadas ao reconhecimento terminológico de uma unidade lexical, conforme frisam Krieger e Finatto (2004, p.133):

> Assim, decidir sobre a admissão de um determinado termo em um dicionário especializado, seja ele um substantivo ou sintagma, implica identificar e atestar sua importância terminológica. Esse é um outro ponto fundamental para todo aquele que se ocupa de Terminologia, de Terminografia e das terminologias. Os princípios adotados para sustentar uma decisão de inclusão, ainda que difiram de um campo para outro, não podem ser arbitrários ou variáveis a todo momento.

Nesse sentido, é importante salientar o quão estão relacionados o embasamento teórico, a sistematicidade organizacional, a familiaridade com a área analisada e a prática terminológica.

Para que se tenha uma aproximação de uma área de conhecimento terminográfico, as normas ISO recomendam criar uma árvore de domínio, ou seja, um diagrama hierárquico contendo termos-chave de uma dada especialidade. Esse organograma auxilia na identificação de inter-relações conceituais de uma área, assim como de outras especialidades conexas. Ademais, contribui para situar um trabalho específico de reconhecimento de termos para um dicionário ou glossário (KRIEGER E FINATTO, 2004, p. 134).

Outro aspecto muito importante é a criação de fichas terminológicas para organizar os repertórios de terminologia, já que nelas podemos encontrar a fonte textual de coleta de um termo, segmentos de texto que mostram a ocorrência do termo em análise, sinônimos, informações sobre as variantes, construções recorrentes que o acompanham, dados do registro, do responsável pela coleta, entre outras informações.

A partir dos dados contidos nas fichas terminológicas, as informações extraídas irão compor o verbete, não sendo necessário repassar todos os dados para o usuário. Além disso, tais informações auxiliarão nos itens de informações como listas, índices, subdivisões temáticas, etc. Cada ficha terminológica reúne, em campos diferentes, todas as características disponíveis referentes a um conceito especializado, como termos, classes gramaticais, contextos, definições, marcas de uso, entre outras. Vale ressaltar que não existe um modelo único de ficha, uma vez que as especificidades dos trabalhos são distintas. O importante é que as informações sejam recuperáveis facilmente e que sejam entendidas por todos da equipe de trabalho, seja ela registrada no papel, ou informatizada.

L'Homme (2006, p.183) salienta algumas recomendações na elaboração de dicionários especializados, que corroboram com as informações que foram mencionadas anteriormente, tais como: (i) levar em consideração as necessidades reais do usuário; (ii) incluir itens altamente especializados e também os menos especializados da área a ser estudada; (iii) pautar as escolhas dos termos em evidências em corpora; (iv) incluir mais dados sobre os termos, como colocações, padrões de valência, imagens, entre outros; (v) contabilizar as diferenças interlinguísticas nos dicionários bilíngues e multilíngues; (vi) descrever as relações entre os termos; (vii) apresentar informações enciclopédicas ou pragmáticas e (viii) estruturar as definições de modo que os componentes conceituais possam ser exibidos.

Para compreendermos um domínio específico é essencial identificar as entidades e as ações básicas desse campo de especialidade, que estão interconectadas por meio de diversos tipos de relações conceituais. Dessa forma, entendemos o conhecimento especializado como uma subdivisão do conhecimento geral, que é composto por esquemas cognitivos flexíveis, que possibilitam a construção de hipóteses e a composição de ideias. Assim, López Rodríguez e outros (2010) assinalam que

> qualquer conceito especializado deve poder encaixar em um sistema formado por *eventos* (atividades, ações, sucessos ou situações), *objetos* (entidades que não são ações nem propriedades) e *relações* que os vinculam entre eles.[21]

Ademais, L'Homme (2006,p. 183) considera que a construção de um modelo de evento para um determinado domínio de conhecimento especializado auxilia o usuário a processar o conteúdo com mais facilidade.

Com base nas informações e orientações descritas nesta seção, nosso intuito é demonstrar no capítulo de metodologia (capítulo 3), quais procedimentos foram adotados, quais foram mais compatíveis com nosso objeto de estudo, a Diciopédia, descrevendo os principais resultados adquiridos.

Após essa breve apresentação dos aportes teóricos que embasam nossa pesquisa, no próximo capítulo, trataremos dos principais dicionários e ferramentas lexicográficas de domínio especializado que foram produzidos com base na Semântica de Frames, apresentando, antes, esta teoria que, além de embasar esses recursos, embasa também a Diciopédia.

---

[21]Cualquier concepto especializado debe poder encajar en un sistema formado por *eventos* (actividades, acciones, sucesos o situaciones), *objetos* (entidades que no son ni acciones ni propiedades) y *relaciones* que los vinculan entre ellos.

2 BASES DE CONHECIMENTO FUNDADAS EM FRAMES

Este capítulo tem por finalidade mostrar como uma teoria linguística, neste caso, a Semântica de Frames, pode auxiliar no desenvolvimento de recursos eletrônicos multilíngues, de modo a moldar as escolhas lexicográficas e as principais informações que constarão nos produtos desenvolvidos. Para isso, começaremos por descrever, brevemente, os aspectos teóricos que compõem a FrameNet.

2.1 Da Semântica de Frames à FrameNet

Nos termos propostos por Fillmore (1982), a Semântica de *Frames* é um programa de pesquisa em semântica empírica que toma emprestada da Inteligência Artificial a ideia de que o conhecimento não deve ser visto como uma coleção de fragmentos simples e desconexos, mas como uma estrutura complexa, denominada frame*,* o qual é definido como “qualquer sistema de conceitos relacionados de tal forma que, para entender um deles, é necessário entender toda a estrutura na qual ele se encaixa” (FILLMORE, 1982, p. 111). Esse campo de estudos procura enquadrar o significado em termos da sua relação com a compreensão geral e não mais nos termos estabelecidos pela semântica das condições de verdade, partindo do pressuposto de que o conhecimento humano é estruturado a partir de um todo compartilhado e não por conceitos isolados, como na Semântica Vericondicional, que, por sua vez, estipulava uma lista de condições necessárias e suficientes, de modo que, para que um determinado lexema se constituísse como uma categoria significativa, traços semânticos deveriam ser capazes de compor sua descrição lexical (FILLMORE, 1985).

Sob esse ponto de vista oposto ao da Semântica Vericondicional, Fillmore (1985) pretende descobrir como um intérprete relaciona a compreensão total de textos linguísticos, a partir do contexto no qual está inserido, ou, como nos afirma Petruck (2008), como ocorre a categorização de expressões específicas em determinadas situações nas quais o falante-ouvinte depreende o significado a partir de sua experiência compartilhada. Fillmore e Atkins (1992) também corroboram essa proposta ao afirmarem que:

> Na Semântica de Frames, o significado de uma palavra só pode ser compreendido com referência a um *background* de experiências, crenças e práticas constituindo uma

> espécie de pré-requisito conceitual para a compreensão do significado (FILLMORE E ATKINS, 1992, p. 76-77).[22]

Segundo Petruck (1996, p.1), um frame é um dispositivo linguístico estrutural cognitivo, no qual há uma indexação das partes às palavras que estão a ele vinculadas a serviço do entendimento. Faber e colaboradores (2005) corroboram este entendimento, ao afirmarem que a noção de frame é uma construção útil não só para definir conceitos, mas também para situá-los dentro de um contexto em que as categorias são relacionadas entre si. De acordo com a Semântica de Frames (FILLMORE, 1982, 1985), o significado das unidades lexicais (ULs) é construído com base na experiência ou conhecimento de *background*, estruturando desse modo como a mente humana armazena informações segundo nosso conhecimento de mundo. Assim, *"*os frames são enquadramentos específicos unificados de conhecimento, ou esquematizações coerentes de experiência*"* (FILLMORE, 1985, p. 223)[23].

Podemos considerar que uma das propostas da Semântica de Frames é a descrição dos significados dentro de uma representação esquemática a partir dos frames, que são considerados uma estrutura cognitiva, dinâmica e flexível, podendo ser aplicados a diversos níveis de categorizações e campos distintos. Neste sentido, há cada vez mais uma interligação entre outras áreas e essa teoria, de modo a possibilitar o surgimento e ampliação de estudos e produtos que envolvam a língua e seu uso, conforme mencionam L'Homme et al. (2016):

> A Semântica de Frames também é cada vez mais citada na Terminologia e em outras áreas com foco em itens lexicais especializados e tem sido usada para descrever termos em diferentes domínios, como o Ambiente, a Lei, Futebol e Biomedicina (Schmidt, 2009; Faber, 2012, entre outros). Frames semânticos são especialmente atraentes na Terminologia já que se presume que há uma conexão entre a estrutura conceitual de áreas especializadas de conhecimento e as unidades linguísticas usadas para transmitir esse conhecimento.[24] (L'HOMME ET AL., 2016).

---

[22]Within frame semantics a word's meaning can be understood only with reference to a structured background of experience, beliefs and practices constituting a kind of conceptual prerequisite for understanding the meaning (FILLMORE & ATKINS, 1992, p. 76-77).

[23]Frames are specific unified frameworks of knowledge, or coherent schematizations of experience (FILLMORE, 1985, p. 223).

[24]Frame semantics is also increasingly cited in terminology and other fields focusing on specialized lexical items and has been used to describe terms in different domains, such as the environment, law, soccer and biomedicine (Schmidt, 2009; Faber, 2012, among others). Semantic frames are especially attractive in terminology since it is assumed that there is a connection between the conceptual structure of specialized fields of knowledge and the linguistic units used to convey this knowledge.

No campo da Lexicografia, a Semântica de Frames tem sido aplicada para auxiliar os lexicógrafos na sistematização de partes de seu trabalho, sendo capaz de exibir as informações de maneira mais útil e amigável para os dicionários especializados. (L'HOMME & CORMIER, 2014, p. 336). Baseando-se nesses fatores, Faber et al. (2006, p. 193), relatam que, nas últimas décadas, tem ocorrido uma aproximação relacionada aos estudos de língua e linguagem, os quais envolvem lexicógrafos, semânticos lexicais e terminológos, tendo em vista uma abordagem mais conceitual na estruturação de dicionários. Um exemplo dessa mudança na lexicografia é o projeto FrameNet.

A FrameNet (doravante FN) é um projeto em lexicografia computacional, iniciado por Charles Fillmore no *International Computer Science Institute* (ICSI), em Berkeley, na Califórnia, e existente desde 1997. Tem como finalidade criar um recurso lexical online para o inglês baseando-se na Semântica de Frames e em evidências em corpora (RUPPENHOFER ET AL., 2016). O escopo do projeto é criar um banco de dados de frames e de seus Elementos de Frame (doravante EF), que são caracterizados por serem os aspectos e componentes de frames individuais, expressos linguisticamente em termos de suas estruturas sintáticas e semânticas, que provavelmente aparecerão nas sentenças de exemplos dos frames aos quais se relacionam, de acordo com suas descrições (FILLMORE E BAKER, 2010, p. 321).

De acordo com os dados disponíveis no *site* oficial do projeto (FRAMENET, 2018), o banco de dados possui mais de 1.200 *frames* e mais de 13 mil unidades lexicais anotadas em mais de 200 mil sentenças em inglês. Leva-se em consideração para a anotação e registro as possibilidades combinatórias que aliam as valências sintáticas e semânticas de cada palavra de acordo com os sentidos estabelecidos (RUPPENHOFER ET AL., 2016). Para essas anotações temos o auxílio de softwares[25] no que tange ao processamento automático de informações, como na demonstração da sumarização das valências, na exibição dos resultados e na apresentação dos grafos, interconectando os elementos de frames a seus frames, bem como os frames entre si, criando assim uma rede semântica. Todos os resultados e informações podem ser consultados gratuitamente no endereço eletrônico https://framenet.icsi.berkeley.edu/.[26] Segundo os dados do site (FRAMENET, 2018), pode ser baixado e usado por alunos, pesquisadores,

[25]FN Desktop

[26]O banco de dados está em constante atualização.

professores para inúmeros propósitos, como, por exemplo, ser um dicionário com exemplos anotados para mostrar o significado e o uso; para quem trabalha com PLN, fornece um banco de dados de treinamento para a rotulação de papéis semânticos, podendo ser usado na extração de informação, na tradução automática e no reconhecimento de eventos; para quem trabalha com a Linguística, pode ser usado como um dicionário de valências sintático-semânticas do inglês.

A partir da estrutura de base da formação do banco de dados da FrameNet do inglês, outras línguas têm expandido esta base de dados, como o espanhol, o alemão, o chinês, o japonês, o francês, o sueco e o português do Brasil, para criarem em sua língua materna uma ferramenta baseada na Semântica de Frames.

Nesses projetos, os frames desenvolvidos pela FrameNet para o inglês são o ponto de partida para as descrições dos léxicos das demais línguas. Conforme já detalhamos sobre o processo de expansão em trabalho anterior (PERON-CORRÊA, 2014, p.23):

> Segundo Lonneker-Rodman (2007), este método de desenvolvimento de recursos lexicais multilíngues é chamado de *Expand* e caracteriza-se por transferir a estrutura de um recurso desenvolvido da língua (A) para outra língua (B). Observadas as especificidades, como, por exemplo, o padrão de lexicalização das duas línguas, adapta-se a estrutura fazendo-se um realinhamento. Assim, ao mesmo tempo em que adequam a base de dados originalmente proposta para o inglês para sua realidade linguística, todos os projetos derivados da FrameNet apresentam o mesmo formato, assim como podem utilizar o mesmo software de anotação lexicográfica, com base no mesmo modelo semântico.[27]

A título de ilustrar o banco de dados e o seu funcionamento, utilizaremos os dados da FrameNet Brasil (doravante FN-Br), que podem ser acessados em http://www.framenetbr.ufjf.br/. Nesta tese, abordaremos, principalmente, os exemplos relacionados ao domínio do Turismo, já que é a área que escolhemos para ilustrar este trabalho. Dessa forma, mostraremos aspectos da organização e categorização conceptual de um domínio, a partir da Semântica de Frames, e adiantaremos a demonstração de como toda essa estrutura pode ser utilizada como princípio organizacional do léxico.

Na Figura 13, temos a definição do frame `Fazer_turismo` e abaixo temos os elementos de frames nucleares (*core*) que compõem este frame, com suas respectivas definições. Estes são assim denominados mediante a sua importância para a

[27]Há outro método para alinhamento de recursos linguísticos, qual seja o *Merger*. Este se caracteriza por criar recursos independentes para cada uma das línguas-alvo, os quais serão posteriormente alinhados (PERON-CORRÊA, 2014, p. 23).

significação e constituição do frame. Também podem aparecer os elementos de frames não-nucleares (*non-core*) denominados periféricos (*peripheral)* e extra-temáticos *(extra-thematic).* Os periféricos se caracterizam por não serem primordiais na modelagem do frame, ou seja, podem aparecer em outros frames: geralmente, são aqueles que preenchem os papéis tradicionalmente descritos como adjuntos de tempo, lugar, maneira. Os extra-temáticos evocam frames separadamente, uma vez que são introduzidos por elementos construcionais, ou seja, expressões que são sintaticamente regidas por um elemento estrutural de suporte, mas transmitem a informação que está fora do frame (FILLMORE, 2012).



Figura 13: Frame `Fazer_turismo`

Fonte: http://www.framenetbr.ufjf.br/ Acessado em 20/07/2017

A FN é uma rede semântica que possibilita a navegação entre os frames interligados, através das relações entre eles, sendo que estas correspondem a mapeamentos de EF a EF. As principais relações são: Herança, Perspectiva, Uso,

Subframe, Precedência, Causativo_de, Incoativo_de e a metarrelação de Veja_também.[28, 29]

A relação de Herança é considerada a mais forte entre os frames, uma vez que o frame-filho é um subtipo do frame-pai, e os EFs do pai estarão presentes no filho. O frame `Locais_naturais`, por exemplo, herda do frame `Local`, uma vez que especifica que a localidade não foi construída pelo homem.

A relação de Perspectiva está relacionada com mais de um ponto de vista sobre um frame neutro, ou seja, há uma perfilação da cena. O frame `Turismo_de_atração` apresenta três perspectivas. No frame `Fazer_turismo`, o EF TURISTA é perfilado para representar a experiência turística, enquanto no de `Atrair_turista`, é o EF ATRAÇÃO que é perfilado.

A relação de Uso ocorre quando temos apenas parte da cena evocada pelo frame filho referindo-se ao frame pai. Trata-se de uma relação mais "genérica", como se houvesse uma pressuposição do frame-filho para o frame-pai. O frame `Viajar` está ligado ao frame `Cenário_do_turismo` por meio dessa relação, já que, prototipicamente, para fazer turismo, o turista viaja.

A relação de Subframe trata de frames mais complexos, no qual temos uma fragmentação do frame mais complexo em frames mais simples. Como exemplo, podemos pensar no frame `Chegada` e seus subframes `Chegada_do_turista_localidade` e `Chegada_do_turista_alojamento`, conforme foi proposto por Gamonal (2013).

A relação de Precedência organiza a relação temporal existente entre os frames, mostrando a disposição sequencial existente entre eles. Assim o *frame* de `Chegada`, precede o de `Estada`, que, por sua vez, precede o de `Partida`.

Por fim, a relação Causativo_de exprime relação de causa, e Incoativo_de se refere a mudanças de estado. Essas são relações semânticas entre frames estativos – `Estar_seco` – e frames incoativos – `Tornar-se_seco` – e causativos – `Causar_secar`. Já a relação Ver_também ajuda o usuário a perceber as especificidades de cada frame, propondo que ele contraste frames muito semelhantes.

---

[28]Na FN de Berkeley também foi implementada a relação de Metáfora. Na FN-Br temos as estruturas qualia (DINIZ DA COSTA, 2017) e as relações de Metonímia e de EF-Frame (GAMONAL, 2017).

[29]Os frames que foram ilustrados em cada tipo de relação já foram mencionados em trabalho anterior, no qual também nosso foco foram alguns frames específicos do Turismo (PERON-CORRÊA, 2014).

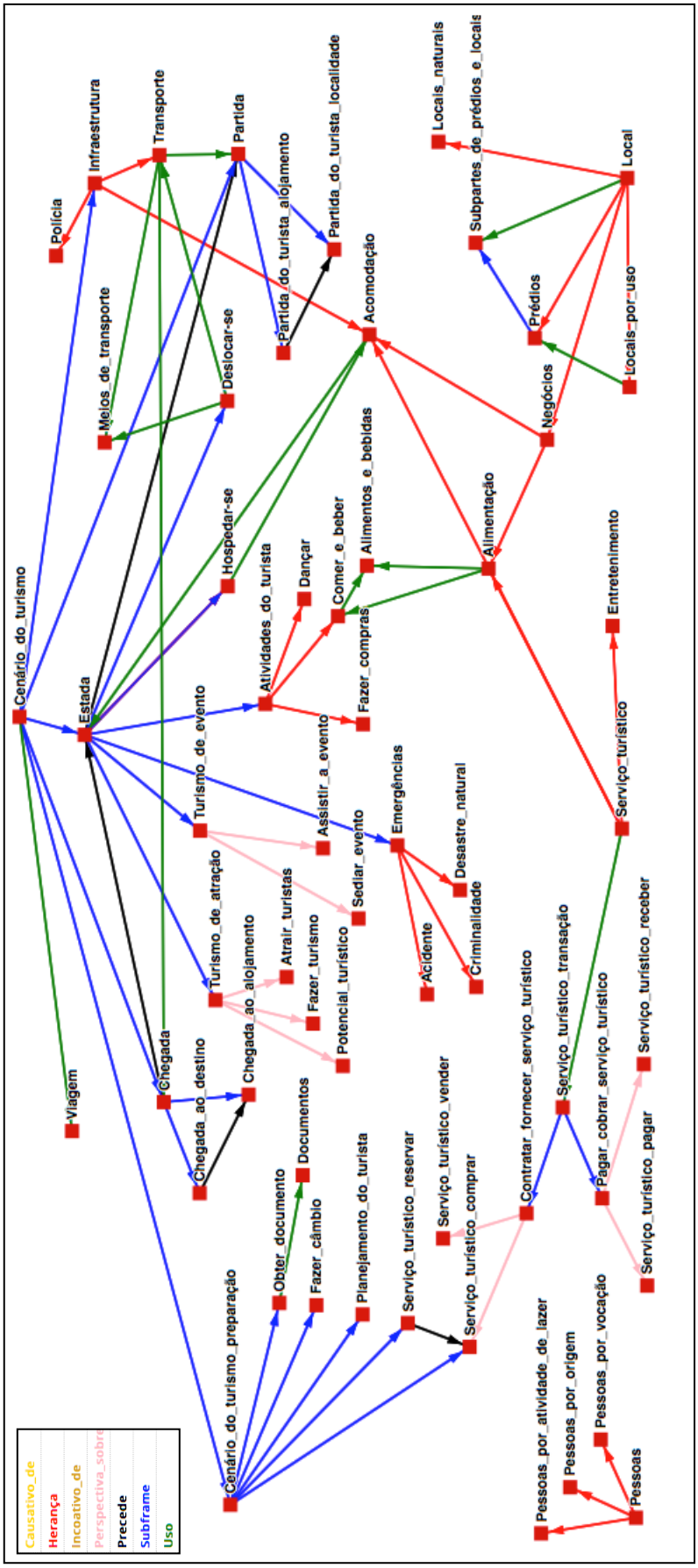


Figura 14: A modelagem do Cenário do Turismo na FrameNet Brasil

Fonte: http://webtool.framenetbr.ufjf.br/ Acessado em 20/07/2017

Na Figura 14, temos a modelagem do Cenário do Turismo e as relações existentes entre os frames. Conforme podemos observar na legenda, cada cor se refere a um tipo de relação que foi explicitado anteriormente.

Uma vez constituída a rede de frames, é possível atestar a ocorrência de ULs que evocam o frame através da anotação de corpus. A anotação na FrameNet apresenta duas metodologias: a lexicográfica e a de texto corrido. Na primeira, selecionamos dos corpora sentenças na qual se instancia uma determinada UL. Nesta etapa, utilizamos a ferramenta WordSketch, presente no Sktech Engine[30], para a extração das sentenças a serem anotadas e visualização dos padrões sintáticos da UL no corpus, fator este que auxilia a intuição linguística quanto ao uso em determinados contextos. Já na anotação de texto corrido, todas as ULs presentes no texto serão anotadas, gerando diversas anotações em uma mesma sentença, com ULs alvo distintas.

Ambas as anotações mencionadas são realizadas através das seguintes camadas: o Elemento de Frame (EF), a Função Gramatical (FG) e o Tipo Sintagmático (TS) do material linguístico que circunda a UL alvo (vide Figura 15). É importante ressaltar que não há a etiquetação das unidades alvo na sentença analisada em relação a sua função gramatical exercida. Conforme já demonstrado em Peron-Corrêa (2014, p. 32), o foco das anotações na FrameNet é:

> (...) registrar as possibilidades de ocorrências de valências sintáticas e semânticas nos diversos sentidos apresentados por cada lexema. Um dos princípios para a anotação lexicográfica é respeitar a localidade sintática (RUPPENHOFFER ET AL., 2010, p. 27), ou seja, os EFs que serão anotados deverão estar situados dentro da projeção máxima nucleada pela UL, salvo exceções como estruturas de alçamento e controle (cf. TORRENT & ELLSWORTH, 2013).

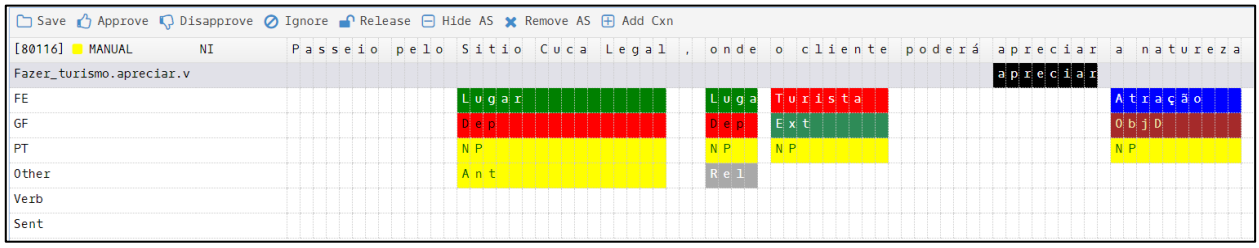


Figura 15: Anotação lexicográfica da UL “apreciar” do Frame `Fazer_turismo`

Fonte: http://webtool.framenetbr.ufjf.br/ Acessado em 20/07/2017

A sentença expressa na Figura 15 mostra a anotação lexicográfica da UL “apreciar”, que evoca o frame `Fazer_turismo`. Temos os seguintes resultados da anotação da

[30]https://www.sketchengine.co.uk/

sentença “*Passeio pelo Sítio Cuca Legal, onde o cliente poderá apreciar a natureza*”: a expressão “Sítio Cuca Legal” e o advérbio “onde” dizem respeito ao EF Lugar; na camada de Função Gramatical[31] (GF) se referem ao Dep e na camada Tipo Sintagmático (PT) é classificado como sintagma nominal (NP). O EF Turista está marcado na expressão “o cliente”, que gramaticalmente exerce a função de externo (Ext) e sua classificação sintagmática também é um sintagma nominal. O EF Atração está se referindo aos termos “a natureza”, que gramaticalmente é um objeto direto e o seu sintagma é nominal. Existem outras camadas, que são usadas em situações muito específicas, como, por exemplo, quando temos um pronome relativo cujo antecedente precisamos marcar, conforme ocorre na sentença acima.

Nesta seção, realizamos um levantamento das principais concepções teóricas concernentes à Semântica de Frames e suas aplicações no que tange ao projeto lexicográfico FrameNet. Ainda que a base teórica seja a mesma entre as FrameNets multilíngues, podemos encontrar diferenças sintáticas, morfológicas ou culturais, de modo que seja necessário realizar um procedimento de reestruturação e até mesmo de realinhamento, para atender às necessidades da língua alvo.

Mediante a estes aspectos, bem como a necessidades oriundas das aplicações computacionais desenvolvidas, a equipe brasileira implementou um novo banco de dados, gerido pela ferramenta FrameNet Brasil WebTool. Na próxima seção, faremos um breve levantamento sobre os principais aspectos inovadores dessa estrutura de dados, com ênfase para o tratamento do Léxico, projeto a que esta tese se vincula.

2.1.1 FrameNet Brasil: Aprimoramento do Banco de Dados

A FN-Br é um desdobramento da FrameNet de Berkeley, que utilizou a estrutura definida para o inglês como ponto de partida para o português do Brasil, fazendo adaptações e alterações, como, por exemplo, incluir novos frames e novas ULs, de acordo com a demanda.

Diante da reestruturação no uso dessa mesma ferramenta para outros idiomas, como realizamos na FN-Br, Boas (2005) menciona que, quanto mais línguas são descritas utilizando a abordagem da FrameNet, proporcionalmente, teremos um

[31]Para maiores informações sobre as funções gramaticais e os tipos sintagmáticos usados na FN-Br conferir Torrent & Ellsworth (2013), e para os que são usados na FN de Berkeley vide Ruppenhofer et al. (2010).

alinhamento de frames semânticos universais, como, por exemplo, temos nos frames de `Comunicação` e `Movimento`, que podem então servir como uma verdadeira interlíngua[32].

Mediante ao desencadeamento de alguns projetos na FN-Br, como um repositório de construções do Português do Brasil (LAVIOLA ET AL., 2017), e recursos multilíngues baseados em frames (TORRENT ET AL., 2014a; PERON-CORRÊA ET AL., 2016), foi necessário desenvolver um novo sistema de anotação e de gestão de banco de dados para a FN-Br, que foi denominado de FrameNet Brasil WebTool.

Esses projetos desenvolvidos na FN-Br estão condicionados aos dois principais recursos que moldam o banco de dados: o FN-Br *Lexicon* e o FN-Br *Constructicon*. Além disso, há também um banco de dados específico trilíngue (português brasileiro, inglês e espanhol) para os domínios do Turismo e Esporte.

Em 2014, com o evento da Copa do Mundo sediado no Brasil, tivemos o lançamento de um dicionário eletrônico trilíngue para os domínios do Turismo, Futebol e Copa do Mundo, através do qual foi demonstrado como os frames poderiam ser usados como representações interlinguísticas, conforme demonstram Torrent et al. (2014a, 2014b), Gamonal (2013) e Peron-Corrêa (2014). Esse recurso multilíngue, Dicionário FrameNet Brasil da Copa do Mundo (http://dicionariodacopa.com.br/), possui 128 frames trilíngues e 1.125 unidades lexicais. Para que fosse gerada uma lista de equivalentes de tradução, foram anotadas 13.000 frases que foram evocadas por ULs dos domínios específicos, principalmente, verbos e substantivos de ação (PERON-CORRÊA ET. AL 2016). Depois dessa experiência, em busca de aperfeiçoar o dicionário e transformá-lo em um guia turístico de bolso, foram necessárias mudanças, que fomentaram a FN-Br WebTool.

Segundo Torrent et al. (no prelo), para o desenvolvimento da FN-Br WebTool, dois princípios foram norteadores: (i) reduzir os requisitos para execução da ferramenta pelo anotador e (ii) superar as restrições que eram impostas pelo FrameNet Desktop,

---

[32] Ao se adotar a concepção de *frames* semânticos como interlíngua, pressupõe-se que eles funcionem como uma ligação, ainda que indireta, entre itens lexicais das diversas línguas. Como cada *frame* pode ser evocado por diversas unidades lexicais, usá-los como interlíngua é uma alternativa para a superação das questões relativas à ausência de paralelismo absoluto entre os idiomas. Outra vantagem do uso de *frames* como interlíngua é a de que eles podem auxiliar no mapeamento dos subsentidos específicos das palavras, quando traduzidas de uma língua para outra, com o intuito de se verificar o processo de equivalência entre elas (BOAS, 2005). Esses aspectos já têm possibilitado outros projetos da FrameNet a identificarem as diferenças linguísticas entre idiomas, analisando-os sob perspectivas de diferentes padrões de lexicalização, como podemos ver no estudo de Ohara et al. (2004), em que se faz uma análise contrastiva dos verbos que indicam percurso de caminho (movimento) na língua japonesa, distinguindo-os em relação à língua inglesa (PERON-CORRÊA,2014, p. 58).

ferramenta desenvolvida pela FrameNet de Berkeley, em especial no que tange ao multilinguismo.

Em relação à execução da ferramenta pelo anotador, tivemos avanços essenciais, como, por exemplo, acessar as funções da WebTool através de um navegador que permita *cookies* e *Javascript*, eliminando a necessidade da instalação de software específico, ampliando dessa forma o acesso de vários lugares. Já em relação ao segundo princípio, implementou-se um banco de dados relacional, atualmente no MySQL, que mantêm a mesma estrutura e conceitos da FrameNet de Berkeley com a finalidade de auxiliar o alinhamento e a migração dos dados (TORRENT ET AL., no prelo).

Neste sentido, um aspecto de extrema importância para o multilinguismo é ter um modelo que possibilite tanto a extensão quanto a integração de outros modelos. Como exemplo, um grande avanço obtido foi o de que, ao invés de termos o trabalho de criar três definições específicas de cada frame para cada língua, ou seja, triplicar a informação, utilizando posteriormente uma relação de tradução para os conectá-los (conforme metodologia utilizada para confecção do Dicionário da Copa utilizando o Desktop), a WebTool criou um registro para cada frame com uma breve descrição e um nome mnemônico. Cada entrada será um registro e terá definições específicas do idioma que o usuário selecionar na interface. Isto para os frames que forem considerados interlinguísticos. Para aqueles em que não há uma correspondência precisa, ou seja, para frames parcialmente equivalentes, será conectado apenas o que for correspondente.

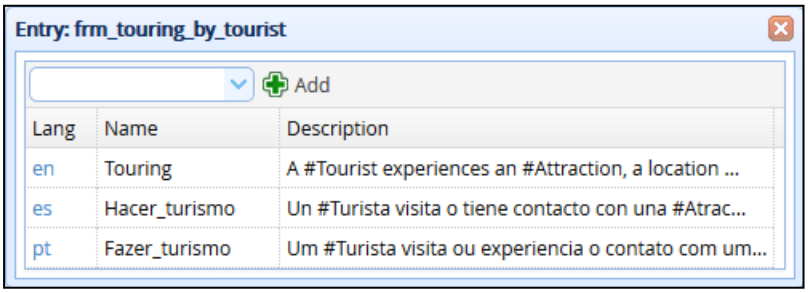


Figura 16: A interface da tabela de entrada para o frame `Touring`

Fonte: Torrent et. al (no prelo) / FN-Br WebTool

Quanto ao processo de anotação, permanecem a lexicográfica e a de texto corrido, conforme mencionamos na seção anterior, com exemplos extraídos da FN-Br WebTool.

O que difere da FrameNet de Berkeley é a anotação de construções[33] que foi criada e que não será abordada nesta tese.

No que diz respeito ao enriquecimento da FN-Br, foram implementadas novas relações: uma entre EFs e frames, uma de substituição metonímica entre EFs (GAMONAL, 2017) e as relações qualia propostas por Pustejovsky (1995). As quatro estruturas qualia são caracterizadas como um sistema de compreensão e inferência de como os seres humanos entendem os objetos e sua relação com o mundo. O quale Formal descreve um dado objeto dentro de seu domínio semântico, caracterizando-o formalmente de acordo com o contexto. O quale Constitutivo expressa o vínculo existente entre um dado objeto e suas partes. O papel Télico está ligado ao propósito ou função de um objeto ou relacionado a um ato que um agente realiza. Por último, temos o Agentivo que determina a origem do objeto, como se deu sua criação, se seu tipo é natural ou artificial.

Os papéis qualia surgiram como forma de modelar diferentes composições possíveis de predicação do contexto de um item lexical. Contudo, as quatro relações (formal, constitutivo, agentivo e télico) estabelecidas por Pustejovsky (1995) são descrições semânticas vagas. Assim, a FN-Br, no âmbito dos trabalhos de mestrado de Vanessa Paiva e de doutorado de Alexandre Diniz da Costa, busca enriquecer as relações qualia tornado-as ternárias, na medida em que passam a ser intermediadas por um frame. As relações ternárias, explicadas de maneira muito resumida, são instituídas a partir de duas ou mais ULs que são ligadas entre si por meio de um quale enriquecido por um frame de background, assim, UL1 estaria relacionada a um dos elementos de frame nucleares, enquanto a UL2 estaria relacionada a outro elemento de frame nuclear do mesmo frame, sendo que quem especifica a relação semântica é o frame. Somente os elementos de frame nucleares (centrais) expressos podem ser recrutados como mediadores de qualia ternários. As relações são representadas de maneira direcional, isto é, são interpretadas como unidirecionais, embora seja possível criar relações inversas (DINIZ DA COSTA, 2017).

Para ilustrar as relações qualia desenvolvidas pela FN-Br, uma UL relacionada ao Turismo é “mapa.n.” Ela evoca o frame `Texto`. Esta UL “mapa.n” está ligada através do quale constitutivo a “papel.n”, uma vez que mapas podem ser feitos de papel. Tem uma

[33]Para maiores informações vide Torrent et al. (2017)

função prototípica de orientar as pessoas, assim, o télico de “mapa.n.” é o verbo “guiar.v”. O agente que produz um mapa é um cartógrafo, dessarte, modelamos “cartógrafo.n” como agente de “mapa.n.” E, finalmente, como um mapa é um tipo de texto, a UL “texto.n” é formal de “mapa.n”, pois distingue o mapa em um domínio maior, conforme mostra a Figura 17.

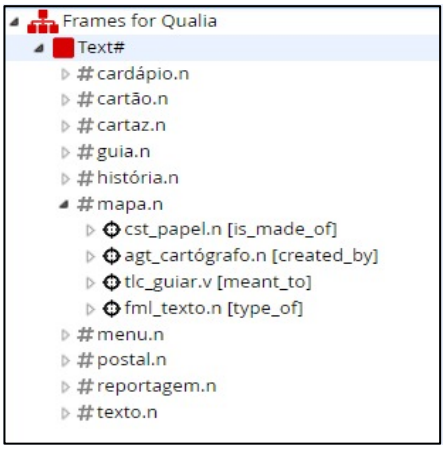


Figura 17: Resultado dos papeis qualia para a UL “mapa.n”.
Fonte: http://webtool.framenetbr.ufjf.br/. Acesso em 26/12/2018

Para conectar essas relações, de acordo com a Figura 18, precisamos relacionar as ULs e o frame de background que especifica cada subtipo de relação. Assim sendo, a relação *Criado por* apresenta como frame de background `Criar_intencionalmente`; a relação *Feito com* tem o frame de background `Ingredientes`; a relação *Tipo de* possuiu o frame de background `Tipo`; a relação *Destinado a* tem o frame de background `Finalidade_do_Utensílio`. O frame aparece para especificar o nível de detalhamento da relação que a princípio, conforme definido por Pustejovsky (1995), pode ser mais genérica.

Quanto às unidades lexicais presentes na Figura 18 temos os seguintes frames evocados: “mapa.n” evoca o frame de `Texto`; “cartógrafo.n” evoca o frame de `Pessoas_por_vocação`; “papel.n” evoca o frame de `Substâncias`; “guiar.v” evoca o frame de `Cotema`.

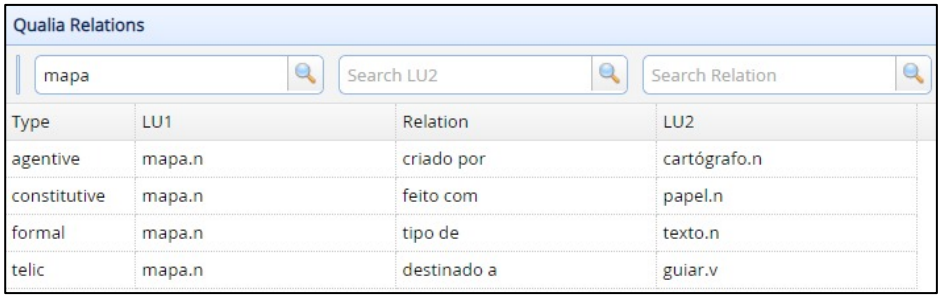


Figura 18: Relações qualia na FrameNet Brasil

Fonte: http://webtool.framenetbr.ufjf.br/. Acesso em 26/12/2018

A Tabela 1 mostra algumas das relações qualia ternárias modeladas até o momento no banco de dados m.knob,[34] com o intuito de dar mais especificidade e granularidade para a descrição semântica dos dados e das relações que compõem as informações prestadas na consulta de um determinado termo.

Quale	Frame	EF / UL1	Especificação	EF / UL2
Agentivo	Afetar intencionalmente	Agente	efeito agentivo	Paciente
	Causação	Ator	causa agentiva	Efeito
	Agir Intencionalmente	Agente	causa agentiva	Ato
	Causação	Efeito	causado por	Causa
	Criação culinária	Alimentos produzidos	criado por	Cozinheiro
	Agir intencionalmente	Entidade criada	criado por	Criador
	Atingir primeiro	Nova ideia	criado por	Pensador
Constitutivo	Pessoas por origem	Pessoa	tem origem	Origem
	Pessoas por religião	Pessoa	é um seguidor de	Religião
	Atributos	Atributo	é atributo constitutivo de	Entidade
	Contendo	Conteúdos	está em	Contêiner
	Incluir	Parte	está incluído no	Total
	Ingredientes	Produto	é feito de	Material
	Filiação	Membro	é membro de	Grupo
	Parte interna	Parte	faz parte de	Todo
	Parte inteira	Parte	faz parte de	Todo
	Parte	Peça	faz parte de	Substância
	Subpartes de construção	Parte	faz parte do edifício	Todo
	Parentesco	Ego	parentesco	Alter
	Residência	Residente	vive em	Localização
Formal	Exemplar	Instância	exemplo de	Tipo
	Tipo	Subtipo	tipo de	Categoria
Télico	Capacidade	Evento	é a capacidade de	Entidade
	Agir intencionalmente	Ato	é a atividade de	Agente
	Personalizado	Comportamento	é o hábito de	Protagonista
	Vício	Vício	é o hábito de	Viciado
	Finalidade da ferramenta	Ferramenta	destinado a	Propósito

Tabela 1: Relações qualia Ternaria no banco de dados m.knob

Fonte: http://webtool.framenetbr.ufjf.br/. Acesso em 26/12/2018

[34] http: // webtool.framenetbr.ufjf.br/index

Conforme demonstrado em trabalho anterior (PERON-CORRÊA, 2014), quando analisamos a valência sintático-semântica de verbos temos mais informações quanto aos seus predicadores. Neste sentido, as anotações existentes nos possibilitam maiores dados e respectivas análises. Quando tratamos de nomes de entidades, as informações de anotações não são tão informativas em sua valência, diferentemente dos verbos que apresentam uma estrutura argumental. Dessa forma, com a implementação das estruturas qualia temos mais dados para nos auxiliar nas análises e na modelagem de recursos lexicográficos no que tange a nomes de entidades (DINIZ DA COSTA, 2017), conforme se verá no capítulo 4.

A implementação da relação entre EF a frame pode ser útil, conforme mencionam Torrent et al. (no prelo) “para modelar os casos em que todas as palavras que evocam o frame pai podem ter como participantes todas as palavras evocando os frames de entidades conectados aos EFs.” [35] Para ilustrar sua utilidade no domínio do Turismo, temos na Figura 19, a definição do frame `Turismo`, no qual temos como um dos EFs nucleares a palavra “turista”, que estará vinculada às ULs do frame `Pessoas_por_atividade_de_lazer`, conforme consta na Figura 20.

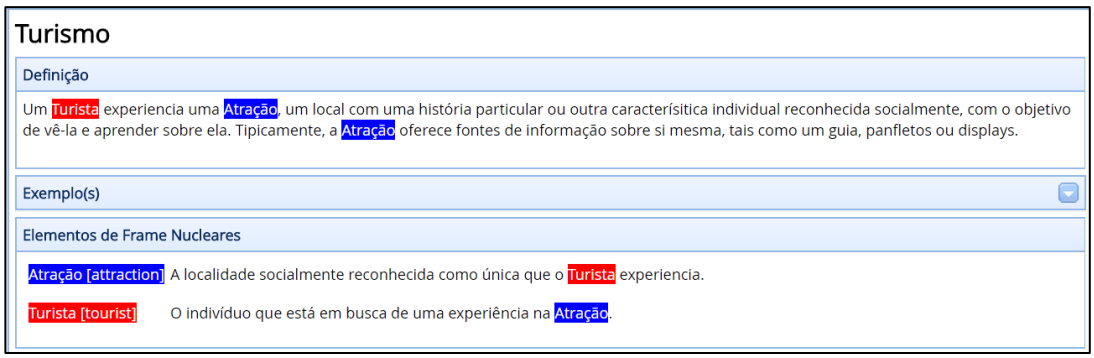


Figura 19: Definição do frame `Turismo`

Fonte: http://webtool.framenetbr.ufjf.br/ Acessado em 20/07/2017

[35]The FE to frame relation is useful to model those cases in which all words evoking the eventive frame may take as participants all the words evoking the entity frames connected to the FEs.

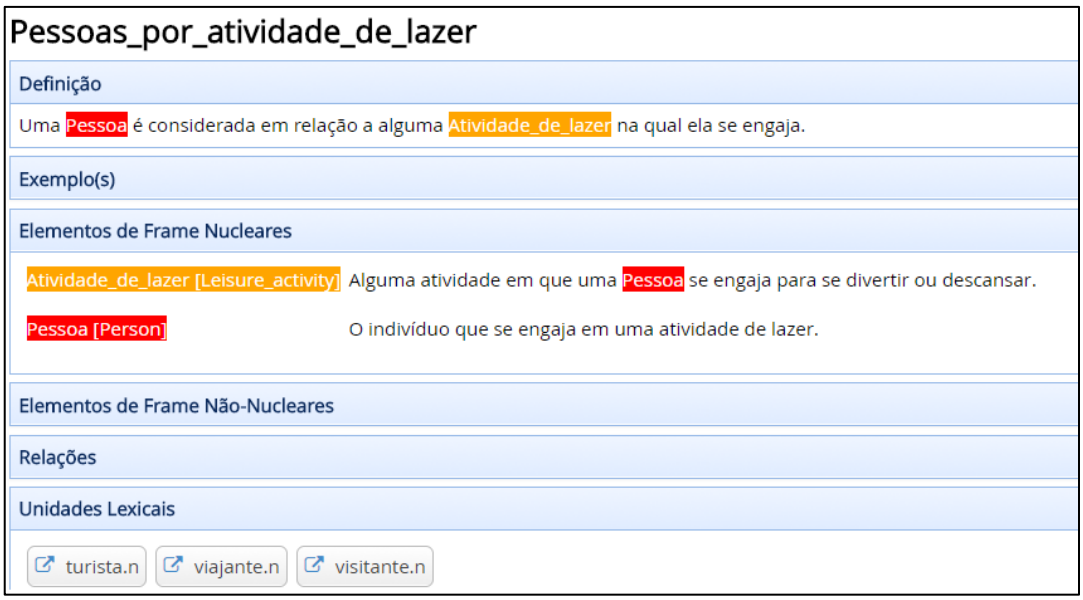


Figura 20: Definição do frame `Pessoas_por_atividade_de_lazer`

Fonte: http://webtool.framenetbr.ufjf.br/ Acessado em 20/07/2017

Todos esses mecanismos explicitados no desenvolvimento da WebTool da FrameNet Brasil tem sido aplicados às ferramentas que estão sendo desenvolvidas, como o m.knob (Multilingual Knowledge Base) (www.mknob.com); o CARMA (Analisador de Construção usando Relações entre Múltiplos AVMs), além de uma nova ferramenta que possibilitará a Extração de Estruturas Linguísticas Relevantes em Gêneros Textuais para fins de Ensino (SIGILIANO E TORRENT, no prelo). Dentre essas ferramentas, nosso foco nesta tese é o aplicativo m.knob, em especial, a sua função Diciopédia. Os dados assinalados nesta seção procuram esclarecer como as informações interlinguísticas são desenvolvidas e até mesmo interconectadas. Com essas modificações pretende-se obter um banco de dados mais denso e mais informativo.

2.2 Desenvolvimento de Recursos de Domínio Específico

Nesta seção, apresentaremos alguns recursos de domínio específico, que foram desenvolvidos a partir da Semântica de Frames e/ou da FrameNet para discutirmos o problema levantado nesta tese, que está relacionado a verificar quais parâmetros metodológicos devem guiar o desenvolvimento de dicionários eletrônicos temáticos multilíngues baseados em frames. Dessa forma, elegemos os dicionários que se ancoram no nosso principal aporte teórico de desenvolvimento lexicográfico, para conhecermos, verificarmos, contrastarmos e discutirmos os principais aspectos que tangenciam tanto a produção lexicográfica quanto seu produto final.

Ao final da seção, elaboramos um quadro comparativo de todos os produtos analisados, conforme os critérios propostos por Atkins (2008, p. 36) para tipificação de dicionários.

2.2.1 EcoLexicon

EcoLexicon (http://manila.ugr.es/visual/) é um recurso terminológico para o domínio do Meio Ambiente, que foi desenvolvido pela equipe de pesquisa LexiCon, da Universidade de Granada (Espanha), sob a direção da professora Pamela Faber. Esse recurso apresenta cada conceito no contexto de um frame especializado, destacando sua relação com outros conceitos. Atualmente, de acordo com os dados do site (http://manila.ugr.es/visual/statistics.html) são 3.468 conceitos com suas respectivas definições, além de 20.368 termos presentes no banco de dados em sete línguas (inglês, espanhol, alemão, francês, holandês, grego moderno e russo). Explorando a convergência de informações linguísticas e gráficas, este recurso pretende atender a diferentes grupos de usuários.

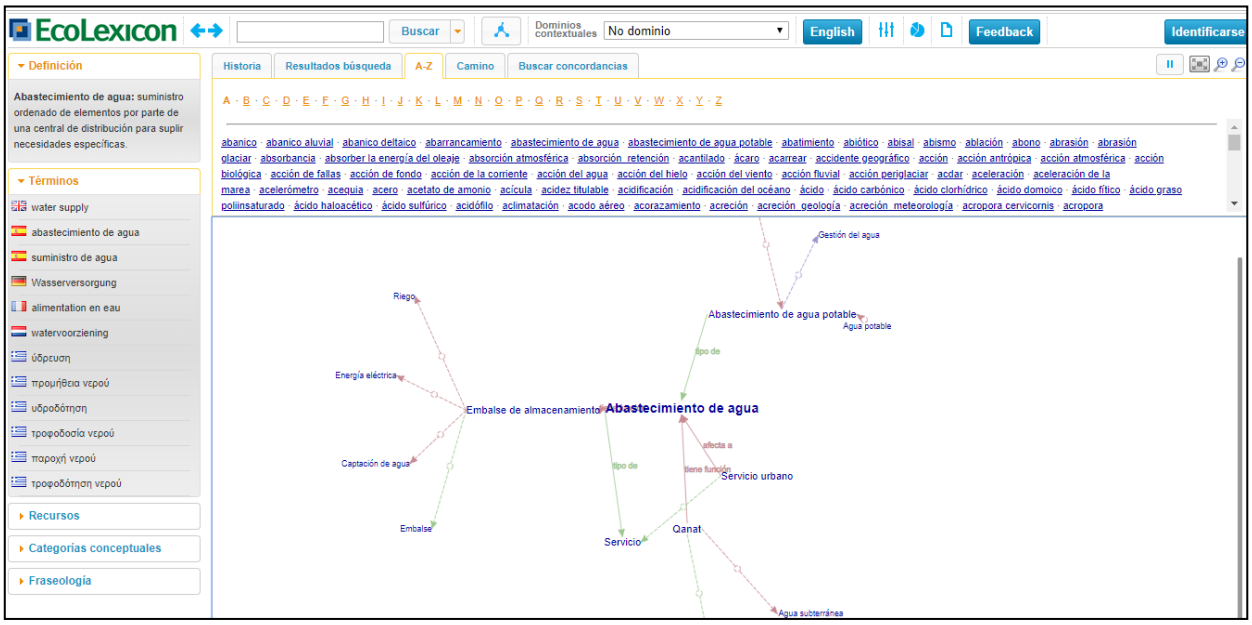


Figura 21: Tela inicial do dicionário EcoLexicon

Fonte: http://ecolexicon.ugr.es/ Acessado em 13/12/2018

Conforme podemos visualizar na Figura 21, do lado esquerdo, temos a definição do termo, assim como sua tradução para as línguas mencionadas. Apresenta também as abas para acessar “recursos”, “categorias conceptuais” e “fraseologia”. As buscas

podem ser realizadas digitando a palavra ou expressão, podem também ser consultadas de acordo com o índice alfabético, ou também podem ser realizadas por meio de “buscar concordância”. Abaixo da busca encontramos a rede de frames que se forma a partir do termo buscado, conforme a imagem mostra a rede relacionada a “abastecimiento de agua”. Cada seta apresenta uma função, por isso, apresenta uma cor diferente.

Segundo Faber e Raúz (2010, p.77-78) o EcoLexicon é uma base de conhecimento multimodal enriquecida com técnicas linguísticas e de representação do conhecimento. Este recurso está ancorado tanto no uso de Ontologias (MEYER, 2001), como no uso da Terminologia baseada em Frames (FABER et al. 2005, 2006, 2007). A importância de tais embasamentos teóricos é defendida por Faber et al (2005), que advogam em favor de:

> (...) uma organização de áreas especializadas baseada em frames em que um frame dinâmico orientado a processos fornece as bases conceituais para a localização das sub-hierarquias de conceitos dentro de um evento de domínio especializado, e a elaboração de um modelo de definição, abrindo assim a porta para uma representação mais adequada dos campos especializados, bem como fornecendo uma maneira melhor de ligar os conceitos dos termos[36].

Assim, mesmo em projetos de linguagem especializada, qualquer conceito deve estar inserido em um frame, o qual inclui suas principais entidades, ações e relações nesse campo de especialidade. Faber e Araúz (2010, p. 79) descrevem os três pilares da Terminologia baseada em frames: (i) caráter organizacional dos termos; (ii) multidimensionalidade dos conceitos e (iii) extração de informação semântica e sintática através de corpora multilíngues.

López Rodríguez et al. (2010) exemplificam o processo de desenvolvimento de dicionários de domínio específico com os projetos MARCOCOSTA e ECOSISTEMA,[37] que foram projetos precursores, que apresentam dois objetivos centrais: estruturar o conhecimento especializado sobre as Ciências Ambientais com o intuito de criar diferentes recursos terminológicos multilíngues e identificar a estrutura conceitual subjacente às Ciências Ambientais, seus principais conceitos, as relações entre os conceitos e sua multidimensionalidade, que pode ser caracterizada pelas variadas

---

[36] (…) a frame-based organization of specialized fields in which a dynamic process-oriented frame provides the conceptual underpinnings for the location of sub-hierarchies of concepts within a specialized domain event, and the elaboration of a definition template, thus opening the door to a more adequate representation of specialized fields as well as supplying a better way of linking terms to concepts.

[37] Esses projetos podem ser acessados no site: http://manila.ugr.es/visual/

possibilidades de organizar os diferentes conceitos, que representam a mesma realidade física dentro de um único sistema (ROGERS, 2004, p.218).

Neste projeto, os conceitos mais primitivos estão organizados em um evento baseado nos frames do EVENTO MEIO AMBIENTAL. Nele encontramos variados processos dinâmicos que se caracterizam por um *AGENTE* (natural ou humano), que incide sobre outra entidade com função de *PACIENTE* (as entidades que formam a costa e a superfície terrestre), no qual produzem um *RESULTADO*, tais aspectos representam a estrutura do núcleo organizacional que abrange a grande maioria dos termos no domínio.

López Rodríguez et al (2010) descrevem que, para encontrar um frame conceptual de referência, é importante que analisemos as definições terminográficas dos conceitos-chave do campo de especialidade em estudo, com a finalidade de identificar os conceitos e as relações conceptuais mais simples que sustentam todo o sistema, tendo em vista que o ser humano constrói o conhecimento especializado a partir dos processos e objetos que são familiares. Para ilustrar, López Rodríguez et al. (2010) mencionam um exemplo em que os componentes (*agente, paciente, resultado*) podem ser reproduzidos na área mencionada. Um AGENTE (ondas) causa um PROCESSO (erosão), que afeta uma entidade geográfica (PACIENTE), que gera como RESULTADO a praia.

Inicialmente, eles utilizaram a metodologia *top-down* no qual se elabora uma lista intuitiva de domínios e subdomínios conceituais. Posteriormente, assinalaram as unidades lexicais ou terminológicas com a metodologia *bottom-up,* ou seja, o terminólogo ou tradutor extrai os termos do texto original e os encaixa em uma rede conceitual, que se relaciona com um evento prototípico representativo do campo de especialidade em questão.

Assim, cada novo conceito se encontra abrigado em um frame relacionado ao EVENTO MEIO AMBIENTAL, que, por sua vez, começará a estabelecer relações com os outros conceitos, já que se trata de uma base de dados relacional. Estas relações são subdivididas em três categorias: específico-genérico, parte-todo e não-hierárquica. Dentro de cada categoria, outros tipos de relações são encontrados, tais como: *tipo de*, *afeta*, *parte de*, *atributo de*, *resultado de*, *localizado na, feito de*, *acima de*, *delimitado por*, *tem lugar em*, *tem a função*, *em frente de*, *faz com que*, *efetuada por*, *a fase de*, entre outros.

Depois desse processo, as definições de cada lexema são elaboradas mediante a análise no corpus, a partir de informações extraídas de dicionários e de imagens. Para

que elas apresentem tanto uma coerência interna (em que se unificam as categorias dos dados que aparecerão em cada entrada) quanto externa (como se interrelacionam entre si para conseguir um todo), extraem-se esquemas categoriais ou *templates* na constituição da definição terminográfica. É importante que esta apresente de maneira explícita o local que o termo ocupa na hierarquia, assim como seus hipônimos e hiperônimos, e a estrutura conceptual subjacente. Os conceitos considerados mais simples foram listados na Figura 22, como ilustram López Rodríguez et al. (2010):

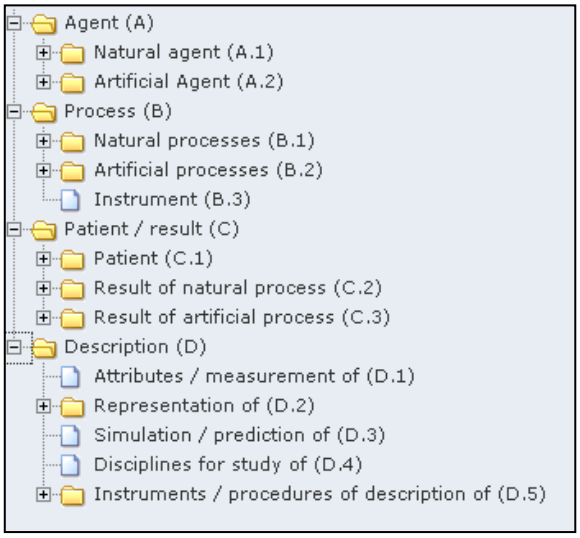


Figura 22: Conceitos básicos do frame conceitual MEIO AMBIENTAL

Fonte: López Rodríguez et al. (2010)

Na elaboração das entradas terminológicas, temos os seguintes passos a serem seguidos: (i) inserção no frame conceitual; (ii) elaboração de definições informativas e coerentes a partir de esquemas categoriais; (iii) implementação de informação gramatical básica; (iv) inclusão de exemplos. Em síntese, López Rodríguez et al. (2010) descrevem a estruturação conceitual do processo

> A estruturação conceitual do domínio das Ciências Ambientais se consegue mediante a elaboração de hierarquias terminográficas, baseadas na extração de informação conceitual de um corpus de textos especializados, da consulta a recursos terminográficos e a especialistas – em concreto, do CEAMA (Centro Andaluz do Meio Ambiente) – e da análise de material multimídia com imagens e animações. O corpus do texto especializado se analisa com a ferramenta de análise do léxico *Wordsmith Tools*, 5, principalmente com as aplicações do *WordList, Concord, KeyWords*y *Viewer*.[38]

[38]La estructuración conceptual del dominio de las Ciencias Ambientales se consigue mediante la elaboración de jerarquías terminográficas, basadas en la extracción de información conceptual de un corpus de textos especializados, de la consulta a recursos terminográficos y a expertos —en concreto, del CEAMA (Centro Andaluz del Medio Ambiente)— y del análisis de material multimedia con imágenes y

Outro aspecto importante da metodologia é a inserção de imagens e definições como representação do conhecimento, segundo López Rodríguez et al. (2010):

> Em nosso projeto incluímos imagens como mecanismo visual de representação do conhecimento em recursos de gestão terminológica, dado que permitem evocar e ativar a estrutura conceptual do domínio de forma mais rápida e complementar a informação semântica, sintática e pragmática codificada textualmente em cada entrada terminológica.[39]

Além de imagens, outros recursos são utilizados no projeto como, vídeos, urls, entre outros.

Na última fase, mapas conceituais são gerados em torno das entradas terminológicas que são relevantes para o EVENTO MEIO AMBIENTAL, e a *posteriori,* novas relações entre os frames podem ser criadas, possibilitando assim a constituição de uma estrutura conceptual superordenada contendo as macrocategorias de um evento, conforme ilustram Faber et al. (2010, apud PRIETO, 2007, 2008), no mapa reproduzido na Figura 23.

---

animaciones. El corpus de textos especializados se analiza con la herramienta de análisis léxico *Wordsmith Tools*,5 principalmente con las aplicaciones *WordList, Concord, KeyWords*y *Viewer*.

[39]En nuestro proyecto incluimos imágenes como mecanismo visual de representación del conocimiento en recursos de gestión terminológica, dado que permiten evocar y activar la estructura conceptual del dominio de forma más rápida y complementaria a la información semántica, sintáctica y pragmática codificada textualmente en cada entrada terminológica.

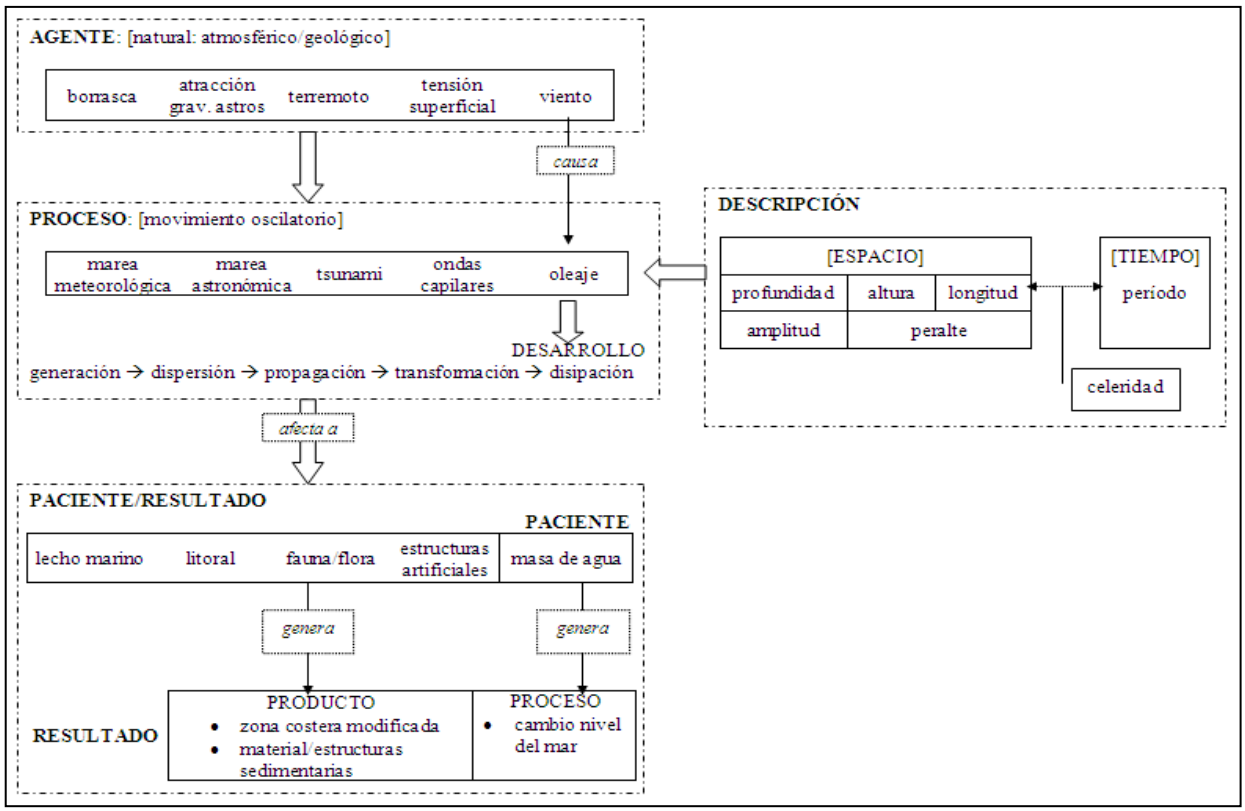


Figura 23: Mapa conceitual do frame de `Marulho`

Fonte: López Rodríguez et al.,2010, apud PRIETO, 2007; 2008.

Com base no frame `Marulho`, temos as categorias *Agente*, *Processo*, e *Resultado*, que apresentam suas unidades terminológicas em cada um dos processos, os quais podem possibilitar o enfoque sob perspectivas distintas, correspondendo inclusive a diferentes fases do processo: origem, início ou fim. Desse modo, segundo Prieto (2007, p. 132), a ativação da representação conceptual do frame de `Marulho`, permite que um mesmo termo possa se referir a distintos eventos, graças ao seu caráter multidimensional, distinguindo-se através da forma como será descrito. Além disso, a configuração em forma de eventos é entendida como uma forma de representação do conhecimento especializado que possibilita maior compreensão dos frames semânticos e das estruturas conceptuais que são subjacentes ao uso das unidades terminológicas.

López Rodríguez et al. (2010) apontam que um dos objetivos da Terminologia baseada em Frames é compreender a representação do conhecimento desde a descrição de como se usa a língua, sendo necessário determinar os termos que representam melhor um conceito, além de inserir as variantes terminológicas na base de dados, já que ela estará relacionada ao termo principal, por meio da semântica. Em suma, a Terminologia baseada em Frames procura representar a multidimensionalidade

dos termos com um vasto repertório de informação conceptual, linguístico, pragmático e visual e o EcoLexicon é uma aplicação desse objetivo.

Na próxima seção, apresentamos outro projeto para o domínio do Meio Ambiente, desenvolvido pela professora Marie Claude L'Homme na Universidade de Montréal.

2.2.2 DicoEnviro

O Dicionário Básico do Meio Ambiente (doravante DicoEnviro[40]), foi desenvolvido pela equipe ÉCLECTIK, do Observatório de Linguística Significado-Texto(OLST), sob a direção da professora Marie Claude L'Homme. Trata-se de um dicionário de domínio especializado, que, além de descrever e enumerar, explica as muitas ligações existentes entre os termos do domínio do Meio Ambiente, no âmbito das línguas francês, inglês e espanhol. De acordo com os dados verificados em outubro de 2014, o DicoEnviro apresenta 70 frames semânticos, contabilizando 744 entradas em francês, 538 em inglês e 158 em espanhol (cada uma corresponde a uma acepção/significado), conforme metodologia expressa em L'Homme e Laneville (2009, p. 9), podendo ser acessado em http://olst.ling.umontreal.ca.

De acordo com L'Homme e Laneville (2009, p. 4), o dicionário é destinado a qualquer usuário que queira aprender sobre a linguagem relacionada ao domínio do Meio Ambiente, incluindo aspectos de gestão de resíduos, energias renováveis e eletrificação dos transportes. Segundo as autoras, há dois níveis de codificação: o primeiro é desprovido da metalinguagem técnica, sendo mais voltado para o usuário que deseja informações sobre termos e suas combinações; enquanto o segundo nível é mais direcionado a linguistas, lexicógrafos ou terminológos que procuram acessar uma linguagem mais técnica para uso de descrição dos termos. Esses são definidos como:

> (...) qualquer unidade lexical cujo significado pode ser associado a uma área específica previamente delineada. Dessa forma, descartamos as unidades lexicais não específicas ao domínio do Meio Ambiente, mesmo se elas são recorrentes em textos especializados (L'HOMME E LANEVILLE, 2009, p. 4).[41]

---

[40]Este dicionário está sempre em constantes atualizações.

[41]Par *terme*, nous entendons toute unité lexicale dont le sens peut être associé à um domain espécialisé préalablement délimité. Ainsi, nous écartons les unités lexicales non spécifiques au domaine de l'environnement, même si eles sont récurrentes dans les textes spécialisés.

Os termos descritos abrangem substantivos (como “planeta” e “mudança”), adjetivos (“climático”, “carbônico”), assim como verbos (“absorver”, “inundar”); advérbios (“localmente”, “mundialmente”) e até mesmo expressões do tipo “efeito estufa” e “força radioativa”.

Algumas das características que classificam o DicoEnviro como original é o fato de fornecer descrições linguísticas no nível léxico-semântico; por se concentrar no funcionamento linguístico dos termos; por se basear em modelos lexicais interligando as unidades lexicais a um frame de domínio específico. Além disso, há uma descrição da estrutura atuacional de cada termo e das ligações paradigmáticas com outros termos do domínio, mostrando também a natureza sintagmática.

A descrição dos termos é baseada na observação destes no corpus especializado, através de um processamento automático que se baseia em um cálculo das especificidades lexicais, através da ferramenta TERMOSTAT (DROUIN, 2003) e também da identificação de equivalentes em inglês (LE SERREC, 2008).

Conforme descrevem L’Homme e Laneville (2009, p. 6), há duas formas de realizarmos a consulta no DicoEnviro. Na primeira versão, encontramos um índice alfabético e uma lista de termos (Figura 24), no qual o usuário clica no item de seu interesse, sendo esta versão atualizada uma vez por dia.

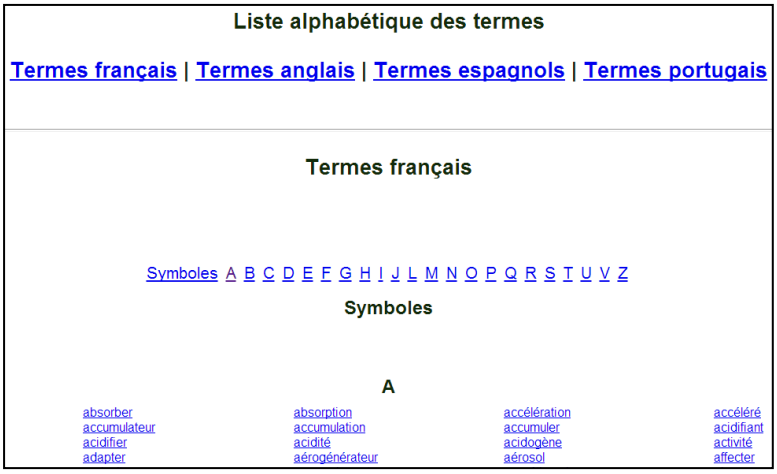


Figura 24: Índice alfabético e lista de termos de consulta no DicoEnviro.

Fonte: L’Homme e Laneville (2009, p. 6).

Na segunda versão, há vários tipos de consulta que podem ser realizadas, conforme mostra a Figura 25.

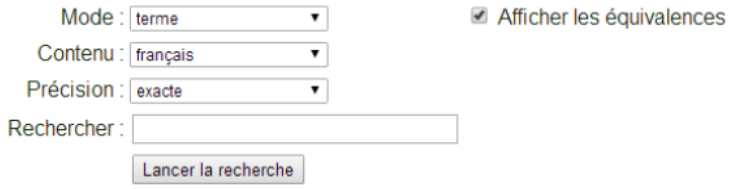


Figura 25: Critérios de consulta no DicoEnviro.
Fonte: L'Homme e Laneville (2009, p. 7).

Dentro do critério "*Modo*" (*mode*), podemos realizar a busca por meio de uma palavra, como "estufa", e ela mostrará expressões em que apareça a palavra, como "efeito estufa", e também pode apresentar sinônimos, variantes e termos relacionados, como "gás carbônico", ou "dióxido de carbono". Podemos digitar um termo, uma expressão ou até mesmo encontrar as traduções de *collocations* ou combinações de palavras frequentemente associadas por meio de uma ligação lexical, ou seja, relações paradigmáticas e sintagmáticas.

No segundo item, "*conteúdo*" (*contenu*), podemos selecionar a(s) língua(s) para encontrar as informações relacionadas a um determinado termo, podendo ser no francês, inglês, espanhol e também na opção multilíngue, na qual aparecem várias línguas. "A opção mostrar os equivalentes permite ao usuário obter equivalências de um termo em outra língua ao iniciar a pesquisa" (L'HOMME E LANEVILLE, 2009, p. 7).

No terceiro item, *precisão (précision)*, podemos aceder aos termos que possuem a mesma ortografia inscrita dentro da janela de pesquisa através do item "exato" (*exacte)*. Por exemplo, se o utilizador escreve "carbono", ele obterá os dados sobre este termo. Contudo, se ele vier a digitar "carb" não obterá resultado algum. Ainda neste item, pode encontrar a busca por *começando com (Débutant par)*, que fornece acesso às palavras que começam com a mesma sequência de caracteres presente na janela de pesquisa. Ainda, permite-se acessar termos que apresentam a sequência de caracteres inscritos no campo da pesquisa através do item "contendo" (*Contenant).*

O DicoEnviro apresenta dez categorias, cujas acepções correspondem a cada item[42] relacionado ao domínio do Meio Ambiente. As categorias presentes em todas as fichas são *Entrada, Informação gramatical, Domínio, Estado/ Status, Estrutura de atuação, Editor e Data da implementação*.

[42]Cada item se refere a cada acepção ou significado neste dicionário.

Na seção *Entrada*, cada uma delas é acompanhada por um número de acepção, independente de o termo apresentar ou não mais de um significado especializado. Quando não há um número de acepção, significa que o termo não faz parte do objeto de estudo no DicoEnviro.

A categoria *Estado/Status* indica se o termo já foi finalizado ou se ainda está em construção. Para ilustrar, se aparecer estado 0, significa que a escrita foi terminada; se aparecer estado 1, indica que o estágio de escrita está muito avançado; se constar estado 2, indica escrita avançada. Tal notação do status indica o nível de progresso para o consulente.

Na seção *Estrutura atuacional e actantes*, encontramos duas partes, sendo a primeira a estrutura atuacional, que lista os actantes semânticos do termo e o seu papel em relação ao que foi descrito. No verbo "queimar", temos dois actantes semânticos em relação ao termo, que são representados por rótulos que descrevem o seu papel em função do termo descrito. Os papéis mais usados neste dicionário são AGENTE, PACIENTE, DESTINO, FONTE e CAUSA. O segundo são os actantes típicos, que correspondem a uma de suas realizações linguísticas, com o propósito de proporcionar uma compreensão melhor do sentido do termo. Também é caracterizado por ser um termo genérico que engloba outras realizações que são utilizadas com mais frequência. Para ilustrar, considere-se o exemplo:

ARMAZENAR $_{1a}$, v. tr.

Papeis atuacionais: DESTINO ~ PACIENTE

Actantes típicos: ecossistema ~ carbono

Na categoria *Contextos*, encontramos como o termo realmente aparece em textos especializados, que formam o corpus. Em algumas fichas, pode aparecer a seção "contextos anotados", que remete principalmente ao trabalho realizado na FrameNet. A anotação na Figura 26 refere-se à UL "reduzir" (ABASSER).

Le *dioxyde de carbone* ainsi emmagasiné, essentiellement dans les couches océaniques supérieures, A ABAISSÉ de *0,1* le *pH* à la *surface* des océans, en raison de sa *nature* acide; on n'a toutefois observé pratiquement aucune modification du pH dans les profondeurs. [5IPCCDYOXYDE 0 ALS 05/06/2008]

La raison est que le *changement* technologique ABAISSE le *coût* des réductions futures par rapport à celui des *réductions* actuelles; il est donc plus rentable de mettre davantage l'accent sur les réductions futures. [8IPCCATTENUATION 0 ALS 05/06/2008]

Figura 26: Anotação da UL *reduzir* no DicoEnviro

Fonte: L'Homme e Laneville (2009, p. 14).

Conforme descrevem L'Homme e Laneville, (2009, p. 14-15), o termo anotado aparece em letra maiúscula; os actantes e circunstâncias são identificados por cores, sendo cada uma delas referente a uma função semântica diferente; os actantes aparecem em negrito e a realização dos actantes em itálico. As informações que sucedem o contexto são a fonte do texto e o status da anotação.

Podemos acessar os frames semânticos que constituem este domínio ao clicarmos no ícone, conforme consta no site do dicionário (http://olst.ling.umontreal.ca). No frame consta a definição da situação mencionada e os elementos obrigatórios (EFs). Os actantes semânticos (participantes 1) e as circunstâncias (participantes 2) são listados em duas colunas, vide Figura 27.

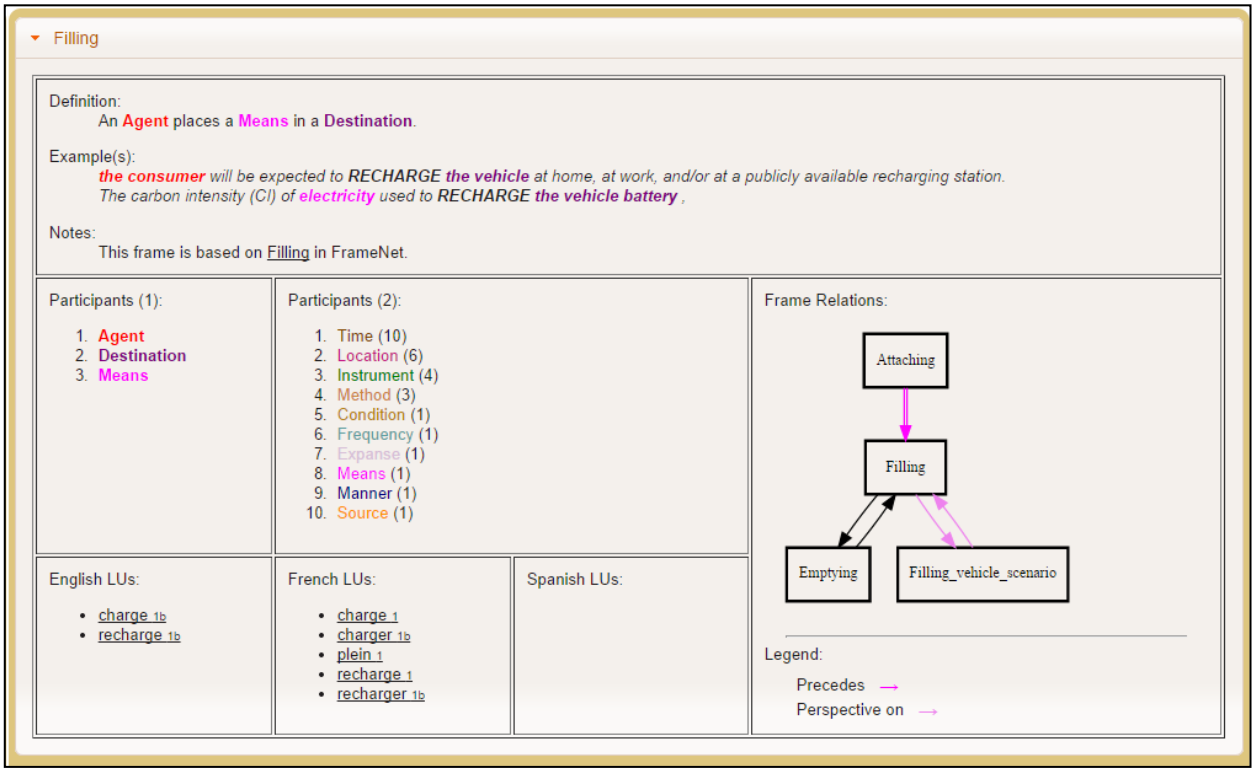


Figura 27: Definição do frame `Charger` (processo de carregamento de uma bateria), no DicoEnviro

Fonte: L'Homme e Laneville, (2009, p. 16)

Para verificar as relações lexicais, o usuário deve clicar na apresentação visual dos links léxico-semânticos que deseja consultar, conforme Figura 28. Nela, os atuantes semânticos estão conectados ao termo AUTOMÓVEL por meio de seta amarela, já os sinônimos por meio de seta verde e os derivados com seta roxa.

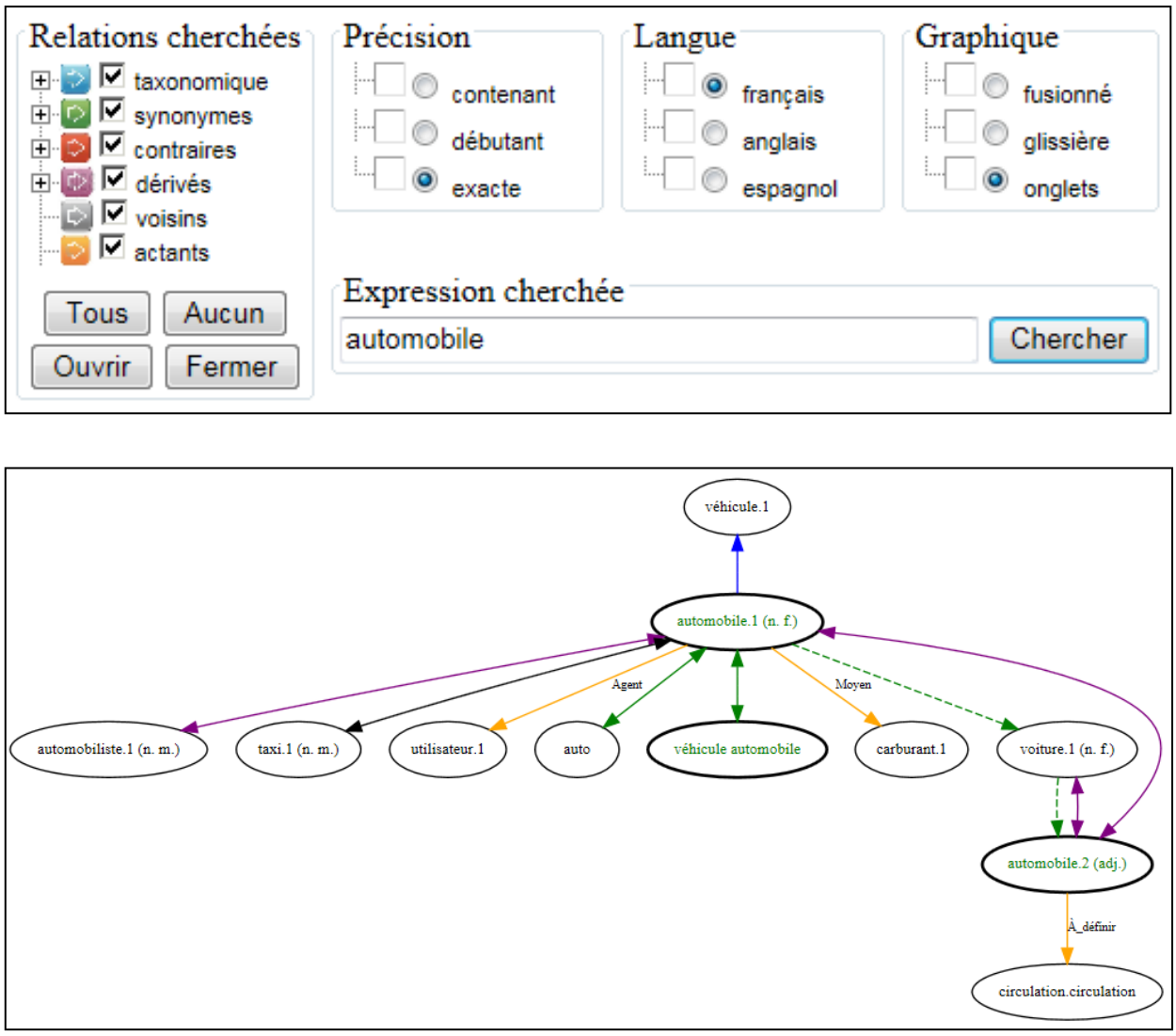
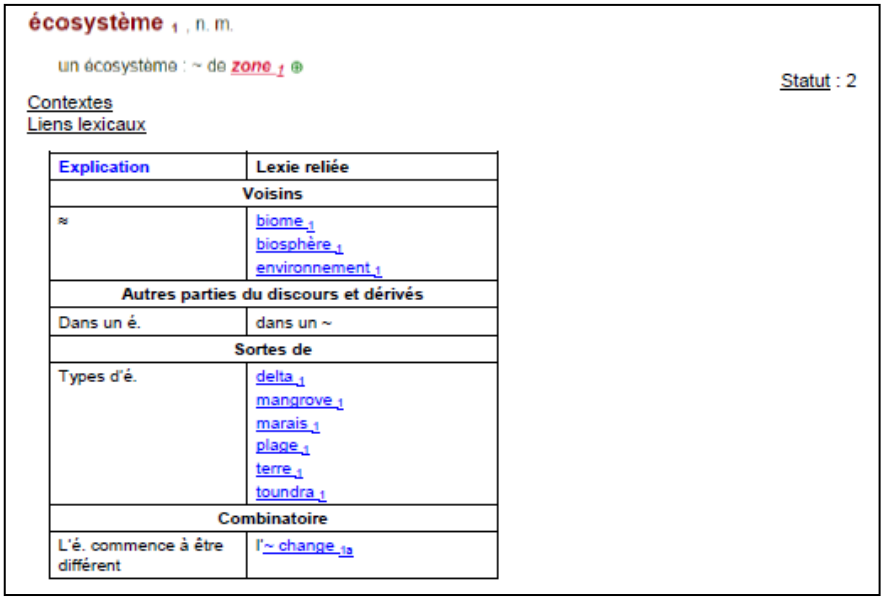


Figura 28: Relações Lexicais no DicoEnviro.

Fonte: L'Homme e Laneville (2009, p. 16)

Por último temos as relações lexicais, paradigmáticas e sintagmáticas, que são exibidas em duas colunas, conforme Figura 29.



écosystème 1 , n. m.

un écosystème : ~ de *zone* 1 ⊕

Statut : 2

Contextes
Liens lexicaux

Explication	Lexie reliée
Voisins	
≈	biome 1 biosphère 1 environnement 1
Autres parties du discours et dérivés	
Dans un é.	dans un ~
Sortes de	
Types d'é.	delta 1 mangrove 1 marais 1 plage 1 terre 1 toundra 1
Combinatoire	
L'é. commence à être différent	l'~ change 1a

Figura 29: Relações Lexicais, Paradigmáticas e Sintagmáticas no DicoEnviro.
Fonte: L'Homme e Laneville (2009, p. 19)

Na coluna da direita, os termos apresentam uma relação semântica com o termo descrito e também são acompanhados por um número de acepção, no qual podemos encontrar um correspondente no dicionário.

Na coluna da esquerda, aparece uma explicação com intuito de facilitar a interpretação das relações semânticas. De acordo com L'Homme e Laneville (2009, p. 19),

> A maioria dos termos que mantém com o termo de entrada uma relação de natureza paradigmática aparece sozinha na coluna da direita. A maioria dos termos que mantém com o termo de entrada uma relação de natureza sintagmática é disposta de forma a comunicar de que maneira eles se combinam com ele.[43]

Os equivalentes de tradução para o termo descrito aparecem em algumas fichas, em inglês e espanhol.

Nessa seção, realizamos um breve panorama sobre o Dicionário especializado para o domínio do Meio Ambiente (DicoEnviro), e, na próxima, trataremos do dicionário para o domínio da Informática e da Internet (DicoInfo).

2.2.3 DicoInfo

[43]La plu part des termes entretenant avec le terme en entrée une relation de nature paradigmatique apparaissent seuls dans la colonne de droite. La plupart des termes entretenant avec le terme en entrée une relation de nature syntagmatique sont placés dans un énoncémontrant de quelle manière ils se combinent avec lui.

Assim, como o DicoEnviro, o DicoInfo (Dicionário fundamental de Informática e Internet), foi desenvolvido pela equipe Eclectik do Observatório de Linguística Significado-Texto (OLST), também coordenado pela professora Marie-Claude L'Homme. Tem como objetivo listar, explicar e interligar os termos básicos do domínio da Informática e da Internet. Na sua versão atual, são 1.900 itens lexicais em francês e1.080 fichas disponíveis no site http://olst.ling.umontreal.ca. A linguagem utilizada é despojada de uma metalinguagem extremamente técnica e é indicado para usuários que tenham interesse nessas áreas, como tradutores, editores, escritores entre outros.

O DicoInfo se concentra nas descrições de unidades lexicais que aparecem com maior frequência nos textos especializados. Os termos que constituem a base de dados podem ser substantivos ("computador", "impressora"), adjetivos ("virtual", "programável"), verbos ("ligar", "infectar"), advérbios ("digitalmente", "manualmente") e também expressões ("sistema operacional", "mapa de memória").

O DicoInfo é apresentado como um produto original, por diferir em relação a outros de domínio especializado devido às teorias que compõem a sua modelagem, como, por exemplo, os modelos lexicais da Teoria Significado-Texto, a Lexicologia Explicativa e Combinatória (MEL'ČUK ET AL., 1984; 1995), Semântica de Frames (FILLMORE, 1982; 1985); e devido à sua aplicação no **DiCo** (POLGUÈRE, 2003) e no **DicoOuèbe** (JOUSSE & POLGUÈRE, 2005). Além disso, descreve a estrutura atuacional de cada termo, apresentando um link paradigmático para conectar outros termos do domínio, como sinônimos, antônimos, entre outros. Outro aspecto de destaque é a anotação sintática e semântica dos termos no contexto, com destaque para os actantes e advérbios presentes.

Segundo L'Homme (2009), os termos são descritos com base no comportamento deles nos textos especializados, que falam sobre internet, computação pessoal, programação, sistemas operacionais, entre outros. A *posteriori*, temos, novamente, um processamento automático na seleção dos termos, propostos por Drouin (2003) e Lemay et al. (2005), que calculam as especificidades lexicais.

A busca no DicoInfo pode ser feita de duas maneiras, pelo índice de ordem alfabética (vide Figura 30), ou por módulo de pesquisa que possibilita diversas opções, como mostraremos a seguir.

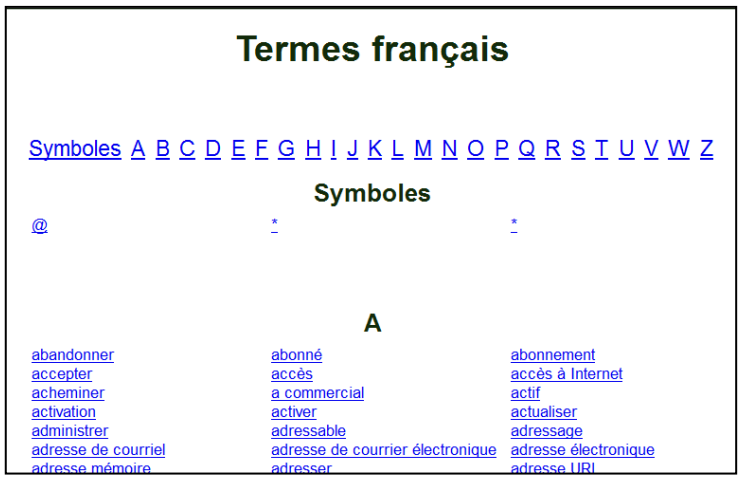


Figura 30: Modo de pesquisa por ordem alfabética no DicoInfo.

Fonte: L'Homme (2009, p. 8)

Através do módulo de pesquisa, ao digitar os caracteres do termo, o sistema apresenta uma série de possibilidades e o usuário deve selecionar uma das opções, conforme mostra a Figura 31.

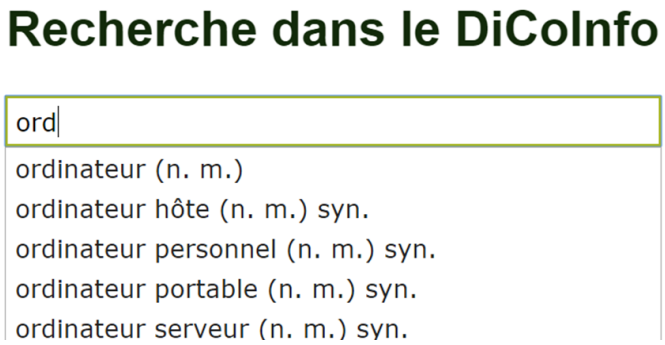


Figura 31: Modo de pesquisa por digitação no DicoInfo.

Fonte: L'Homme (2009, p. 8)

Após a consulta, os resultados são exibidos, constando também os sinônimos e as possíveis variantes (vide Figura 32).



Figura 32: Resultado de pesquisa do termo "web" no DicoInfo.

Fonte: L'Homme, (2009, p. 9)

Podemos realizar a busca dos termos em francês, inglês, espanhol, ou selecionar o campo multilíngue, o qual fornece as opções de idioma mencionadas anteriormente, assim como solicitar a busca por equivalentes de tradução ao iniciar a pesquisa, conforme vemos na Figura 33.

Outros aspectos presentes como opção de busca na ferramenta podem ser visualizados na Figura 33, do lado direito, na qual encontramos, além da seleção do idioma, a partir do item “modo” (*mode*), outros parâmetros de pesquisa, como por “palavra” (*mot*), termo (*terme*), relação lexical (*lien lexical*) e expressão (*expression*). Quanto ao item “Precisão” (*précision*), que estabelece a correspondência da busca do usuário com o conteúdo dos formulários, as opções são as mesmas existentes para o DicoEnviro.

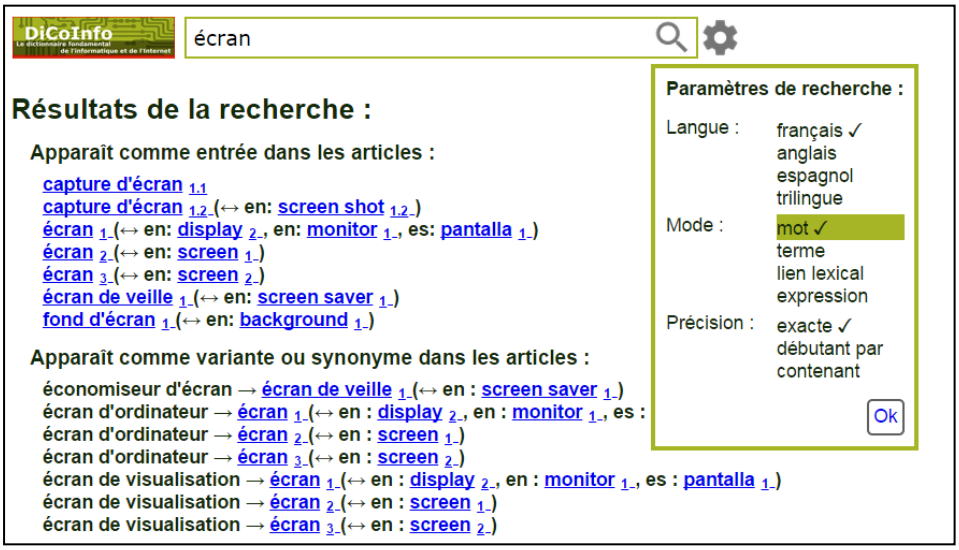


Figura 33:Resultado de pesquisa do termo “écran” no DicoInfo.

Fonte: L’Homme (2009, p. 10)

Feita a pesquisa, encontraremos na entrada a classe gramatical do termo, o gênero, as variantes e os equivalentes de tradução em inglês e espanhol, conforme podemos observar na Figura 34.

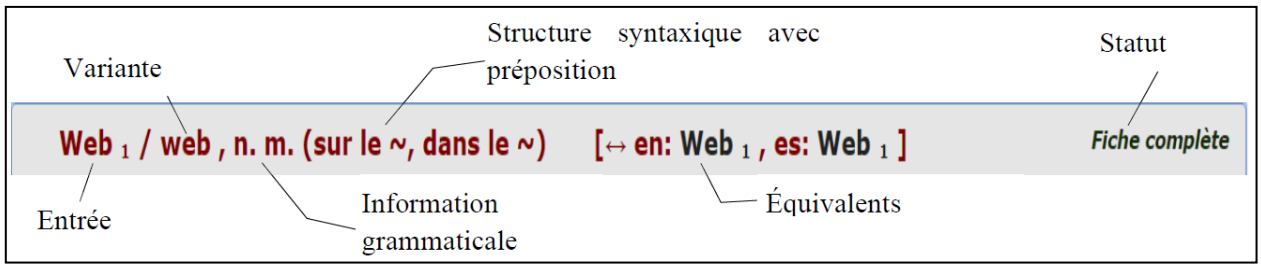


Figura 34: Entrada do termo “web” no DicoInfo.

Fonte: L’Homme, (2009, p. 13)

Cada entrada que pertence ao domínio especializado da Informática ou da Internet elencada no DicoInfo apresentará um número de acepção, mesmo que ela apresente apenas um significado especializado. Se apresentar mais de uma variante, elas serão separadas pelo símbolo da barra [/], como nos exemplos: *Internet 1 / internet ; antivírus1/ anti-vírus.* Alguns termos podem vir distinguidos por letras diferentes, como no exemplo da Figura 35.

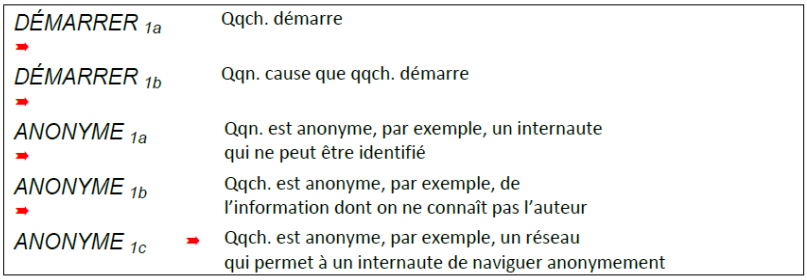


Figura 35: Entrada com acepções marcadas por letras diferentes no DicoInfo.

Fonte: L’Homme (2009, p. 14)

Na busca por termos concretos, esses serão ilustrados por uma imagem (vide Figura 36), que está associada por meio de um hiperlink para acesso. Ao passar o cursor sobre a imagem, aparece o nome do objeto.



Figura 36: Exemplo de um termo associado a uma imagem no DicoInfo.

Fonte: L'Homme, (2009, p. 16)

A exibição das guias, nas quais se encontram as categorias, será descrita na ordem em que são apresentadas na interface. No primeiro caso, temos a possibilidade de visualizar a definição e a estrutura atuacional – vide Figura 37.

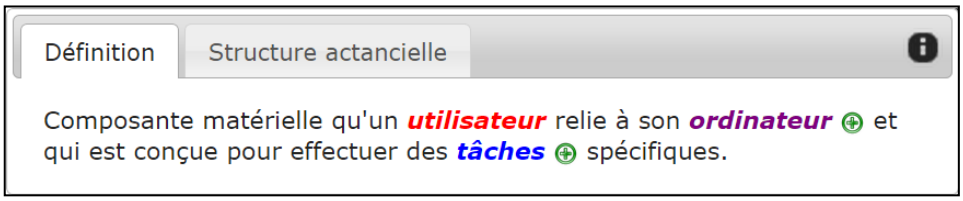


Figura 37: Definição e estrutura atuacional no DicoInfo.

Fonte: L'Homme, (2009, p. 16)

A estrutura atuacional tem por finalidade apresentar os actantes semânticos, que são descritos por meio de papéis (os rótulos mostrados na Figura 38) em relação ao termo pesquisado, de acordo com suas realizações linguísticas (seguindo a mesma estrutura apresentada no DicoEnviro). Os rótulos mais usados neste dicionário são os de AGENTE, PACIENTE, ORIGEM, LOCAL,DESTINO e INSTRUMENTO.[44]



Figura 38: Ilustração da estrutura actancial no DicoInfo.

Fonte: L'Homme (2009, p. 17)

[44]Para maiores informações e detalhes vide L'Homme (2009).

Outra categoria (guia) que consta na interface são as opções de *sinônimos*, *contrários* e *ver também* (nesta opção aparecem os termos quase sinônimos, ou seja, que mantêm uma proximidade, por exemplo, os termos “diálogo” / “discussão”) para o termo pesquisado, conforme mostra a Figura 39.

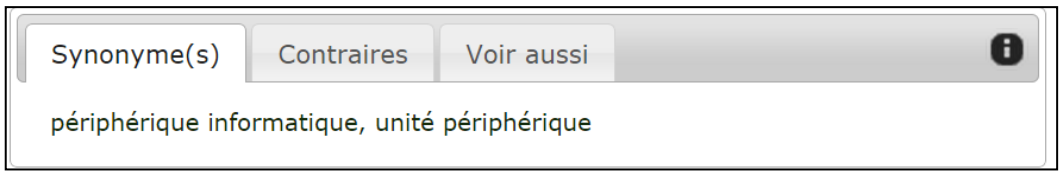


Figura 39: Ilustração da estrutura atuancial no DicoInfo.

Fonte: L’Homme (2009, p. 17)

Em outra categoria, temos os contextos, contextos anotados e resumos. Os contextos servem para ilustrar como os termos são empregados de acordo com o seu uso nos textos. Nos contextos anotados (Figura 40), temos algumas informações linguísticas extras através das etiquetas, como os papéis atuacionais assim como a anotação dos rótulos semânticos dos advérbios.

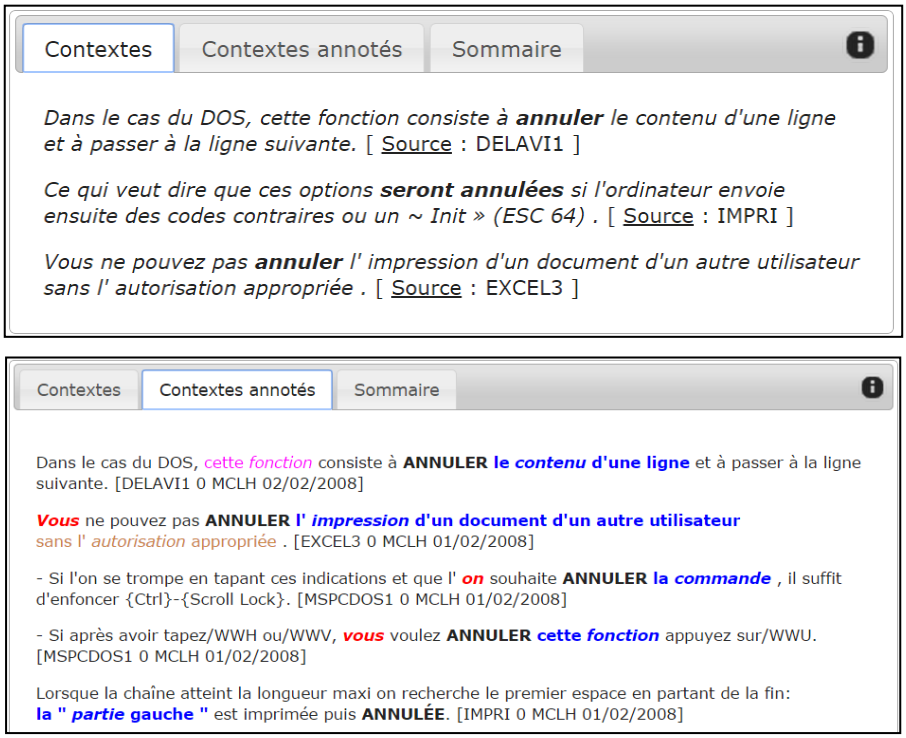


Figura 40: Ilustrações de contexto e contexto anotado no DicoInfo.

Fonte: L’Homme (2009, p. 21, 23)

No contexto anotado, a unidade está escrita em letra maiúscula; temos a anotação para o termo sujeito; os actantes e os advérbios são etiquetados através das cores, de acordo com sua função semântica e os actantes estão marcados em itálico, conforme metodologia descrita para o DicoEnviro.

A próxima guia mostra as relações paradigmáticas, com exceção da sinonímia, que já foi mostrada. Tais relações descrevem a ligação existente entre os termos do domínio, como mostra a Figura 41.

Explication	Terme relié
Dont l'utilisation est facile et agréable	~ convivial 1
	convivialité 1 d'un ~
Un n. qui a fait l'objet d'une préparation en vue d'une utilisation qui convient à l' ***utilisateur***	~ personnalisé 1
Un n. qui peut être préparé en vue d'une utilisation qui convient à l' ***utilisateur***	~ personnalisable 1

Figura 41: Ilustração das relações paradigmáticas para o termo “navegador” no DicoInfo.
Fonte: L’Homme (2009, p. 24)

Conforme demonstrado na Figura 41, na coluna direita, temos os termos que possuem uma ligação com o termo “navegador”. Os termos que estão conectados apresentam um hiperlink que pode gerar uma nova consulta no dicionário. Já na coluna da esquerda, temos o nome mais popular dos recursos lexicais, que estão ocultos.

Também há fichas terminológicas, ou seja, com todos esses dados que descrevemos até aqui no inglês, sendo 850 termos descritos para esta língua, e a versão espanhola ainda em desenvolvimento. Ambas seguem os mesmos critérios descritos. Há extensão desse projeto para outras línguas, como coreano e árabe, que ainda estão em andamento.

Podemos observar que tanto os critérios metodológicos quanto os teóricos são iguais para o DicoEnviro e o DicoInfo, já que foram projetados e concluídos sob uma mesma orientação. Na próxima seção, conheceremos o Kicktionary.

2.2.4 Kicktionary

O Kicktionary é um dicionário eletrônico multilíngue (alemão – inglês – francês) da língua do futebol, que apresenta cerca de 1.900 unidades lexicais distribuídas entre

substantivos, verbos, adjetivos e expressões idiomáticas, nas três línguas mencionadas. Segundo dados do site (http://www.kicktionary.de/index.html), o dicionário foi desenvolvido entre setembro de 2005 e julho de 2006 por Thomas Schmidt.

Segundo Schmidt (2007; 2008), a motivação maior do projeto era construir recursos lexicais baseados em teorias linguísticas, tais como Semântica Lexical, Linguística de Corpus, Tecnologias de Hipertexto e Hipermídia, assim como de Processamento Computacional de Linguagem, que fossem melhores ou contribuíssem de uma forma diferente da dos dicionários de papel tradicionais. Assim, a base teórica desse projeto está ancorada na Semântica de Frames (FILLMORE, 1982; 1985) e na metodologia específica do projeto FrameNet (FILLMORE ET AL., 2003; RUPPENHOFER ET AL. 2016; BOAS, 2009) para auxiliar na construção do banco de dados baseados em frames semânticos; nas pesquisas de Seelbach (2001; 2003) e Gaston Gross (2002), que descrevem sobre o tratamento lexicográfico da linguagem do futebol; nos aspectos linguísticos relacionados às relações lexicais, como sinonímia, hiponímia, equivalentes de tradução, entre outros, que foram modelados na WordNet (FELLBAUM, 1998) e que serviram de base para a modelagem do domínio do Futebol, e, por último, nas sete teses formuladas por Storrer (2001), sobre o uso de hipertexto em lexicografia.

No banco de dados há, aproximadamente, de uma a dez frases anotadas para cada unidade lexical, as quais foram extraídas de um corpus de notícias sobre futebol. A Tabela 2 mostra a quantidade de unidades lexicais anotadas em cada língua (alemão, inglês e francês) e as classes morfológicas mais presentes.

	alemão	Inglês	francês	Todos
Unidades lexicais	792	599	535	1926
Substantivos	451	318	290	1059
Verbos	305	248	201	754
De outros	36	33	44	113
Exemplos	3551	2374	2239	8164
Exemplos de línguas faladas	126	0	0	126
Exemplo / LU	4.48	3.96	4.19	4.24
FOT anotadas	5731	3882	3647	13260
Suportes anotados	554	293	340	1187

Tabela 2: Tabela dos dados anotados do Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Na constituição da estrutura semântica e argumental, as unidades lexicais são organizadas e agrupadas em cerca de cem frames. Há também os synsets (conjuntos

de sinônimos), que são construídos através de uma série de hierarquias conceituais por meio de relações lexicais como hiperonímia, hiponímia, holonímia, meronímia e troponímia.

O corpus utilizado para a extração de sentenças foi retirado do site da UEFA (www.uefa.com), totalizando cerca de duzentas mil palavras. Outro dado importante é que o corpus central é parcialmente paralelo, ou seja, são majoritariamente traduções diretas um do outro para vários idiomas. A anotação lexicográfica segue o modelo da FrameNet[45], conforme podemos observar na sentença (1):

(1) [Yattara] PLAYER_WITH_BALL **evitou** [seu marcador] OPPONENT_PLAYER e disparou de um ângulo agudo.

A unidade lexical que está sendo anotada na sentença (1) é "evitou", por isso está destacada em negrito. O frame evocado é "bater" (*beat*), e o primeiro nome entre colchetes está sendo anotado como EF "Jogador com a bola", e a expressão "seu marcador" está anotada como EF "jogador oponente".

Segundo Schmidt (2007), uma cena corresponde ao conhecimento que o falante tem e pode ativar sobre uma certa situação prototípica. Neste exemplo mencionado, temos dois jogadores (atores) que participam de uma situação individual com um objeto (a bola) em um determinado local (campo), onde ocorre o evento. No Kicktionary, são 16 cenas, sendo que a maioria delas assume a forma de um texto curto, o qual é complementado por um diagrama esquemático, podendo ser uma foto ou desenho que descreva a cena. Para ilustrar, temos um diagrama que representa parte da descrição da cena "passe" (*pass*), reproduzido na Figura 42.

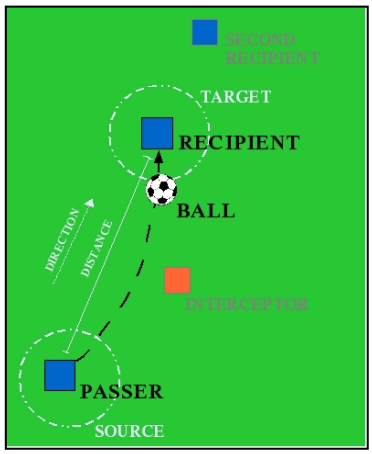


Figura 42: Descrição de parte da cena "passe" no Kicktionary.

Fonte: http://www.kicktionary.de/index.html (acessado em 27 de julho de 2017).

[45] Para maiores informações sobre anotação lexicográfica na FrameNet consultar RUPPENHOFER ET AL. 2016.

Segundo os dados do dicionário, o cenário do “Passe” é centrado em torno do evento de um jogador transferindo a bola para um colega de equipe. Os principais protagonistas do cenário são o jogador que faz o passe e o destinatário. Usando uma parte de seu corpo, o jogador que faz o passe dirige a bola para o destinatário. A bola se move em uma determinada direção a partir da localização da fonte no campo ao longo de um caminho para uma localização de destino, cobrindo uma certa distância.

As unidades lexicais que descrevem a passagem da bola do ponto de vista do jogador que faz o passe fazem parte do frame `Passar`. Já para as ULs que descrevem o passe cujo destinatário é o goleiro do time do jogador que fez o passe o frame é `Passar_de_volta`.

Schmidt (2007, p.191) faz uma distinção entre a definição de frame e cena:

> O frame é uma entidade estrutural usada para agrupar expressões linguísticas que compartilham uma perspectiva comum em uma determinada cena conceptual. Uma cena, nesse sentido, é uma construção superordenada a um frame. É definida em termos de pedaços de conhecimento abstrato, possivelmente não-linguístico, enquanto a noção subordinada de um frame se refere às propriedades dos meios linguísticos concretos de expressar esse tipo de conhecimento.[46]

Na Figura 43, temos uma visão do esquema da estrutura do banco de dados do Kicktionary, que mostra a interligação entre as cenas, os frames, as unidades lexicais e os *synsets*, que estão interligados às hierarquias conceptuais.

[46] “(...)a frame is a structural entity used to group linguistic expressions which share a common perspective on a given conceptual scene. A scene, in that sense, is a superordinate constructo to a frame. It is defined in terms of pieces of abstract, possibly non-linguistic, knowledge, where as the subordinate notion of a frame is concerned with the properties of concrete linguistic means of expressing this kind of knowledge.

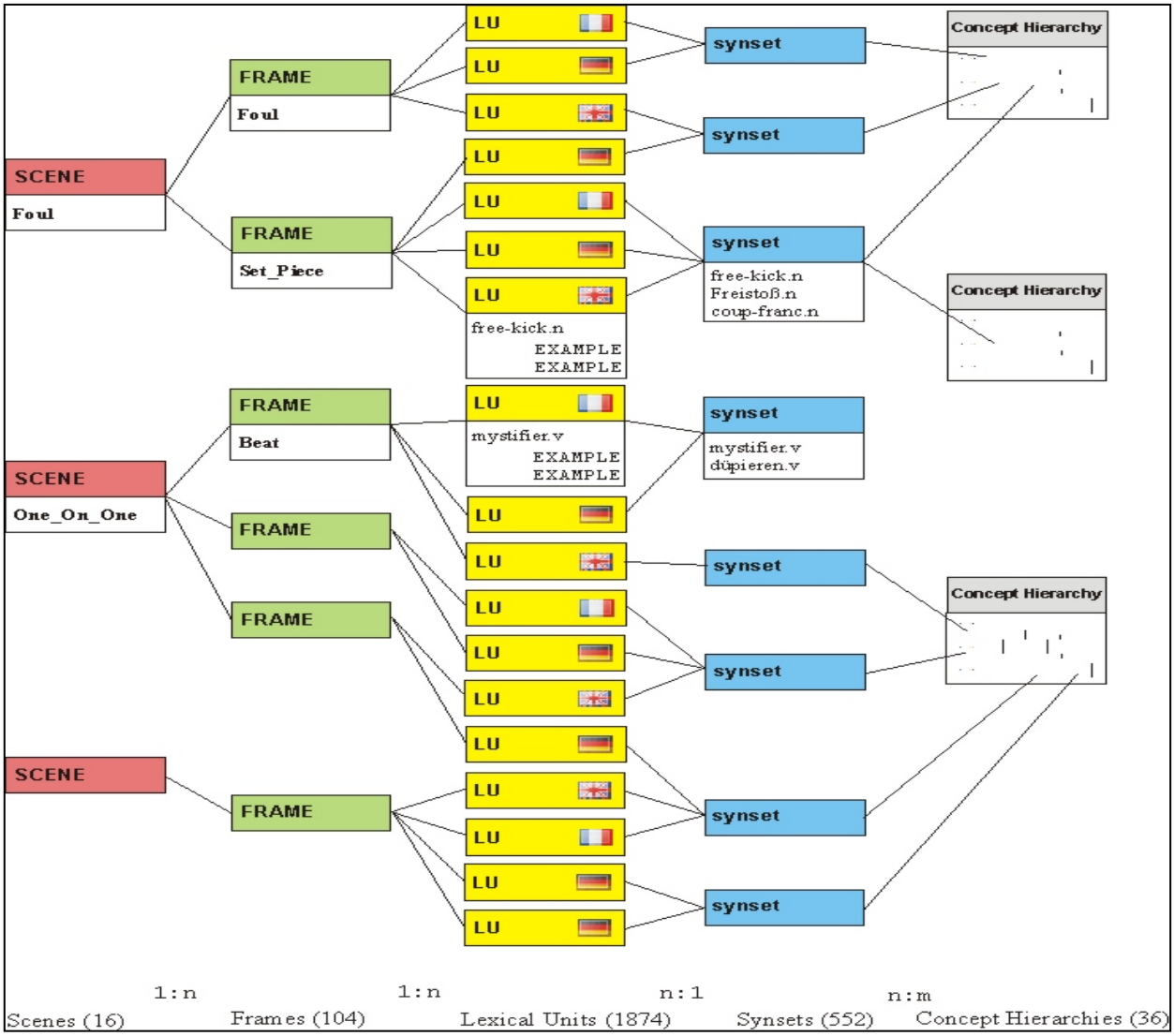


Figura 43: Esquema da estrutura do banco de dados do Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Conforme podemos observar na Figura 43, o banco de dados interliga de diversas formas os aspectos linguísticos que mencionamos, representando-os também nos idiomas inglês, francês e alemão. São dezesseis cenas, cento e quatro frames, mil oitocentos e setenta e quatro unidades lexicais e trinta e seis hierarquias conceptuais. Segundo Schmidt (2007, p. 193), ele utilizou

> a noção de synset não só para agrupar unidades léxicas sinônimas dentro de uma língua (por exemplo, *pênalti* e *penalidade máxima*), mas também para emparelhar uma unidade lexical em um idioma com um potencial equivalente de tradução em outro idioma (p. ex., en. *spot-kick*, de. *Strafstoß* e fr. *coup de pied de réparation*).[47]

[47]The notion of a synset not only to group synonymous lexical units within a language (e.g. *penalty* and *spot-kick*), but also to pair a given lexical unit in one language with a potential translation equivalent in another language (e.g. en. *spot-kick*, de. *Strafstoß* and fr. *coup de pied de réparation*).

Há três relações semânticas diferentes que são modeladas no banco de dados. A primeira é a relação Hiperonímia / Hiponímia, na qual uma é a subclasse mais específica da outra. A segunda é a Holonímia / Meronímia, na qual uma parte constituinte é parte da outra. E, por último, temos a relação de Troponímia, na qual um dos verbos expressa uma forma específica de elaboração de outro, ou seja, se *X for Y de alguma forma*.

As formas de busca presentes no dicionário, conforme podemos constatar na Figura 44, são através de unidades lexicais, cenas, hierarquias conceptuais, exemplos de falas, textos paralelos.

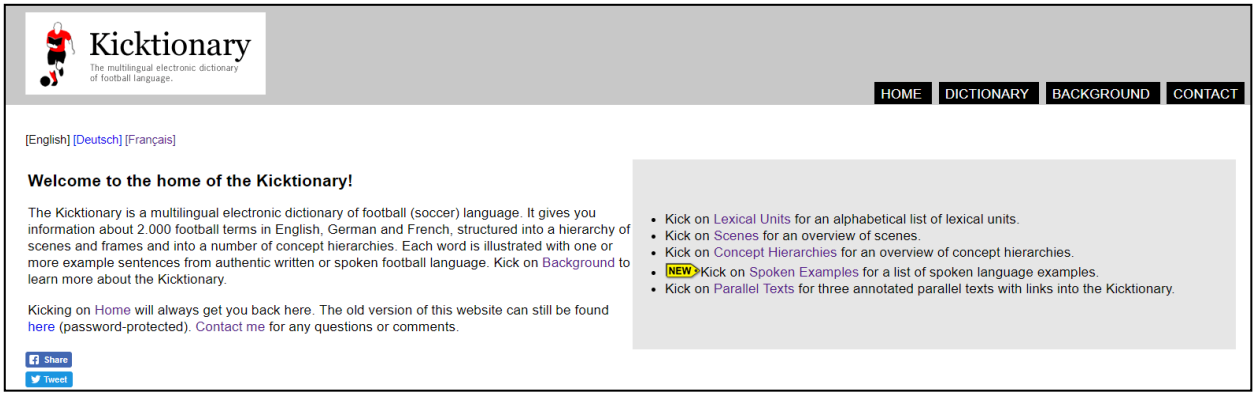


Figura 44: Tela inicial do Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Ao clicarmos sobre a busca por meio das unidades lexicais, é possível digitar a UL que se busca ou encontrá-la em ordem alfabética, como podemos ver na Figura 45.

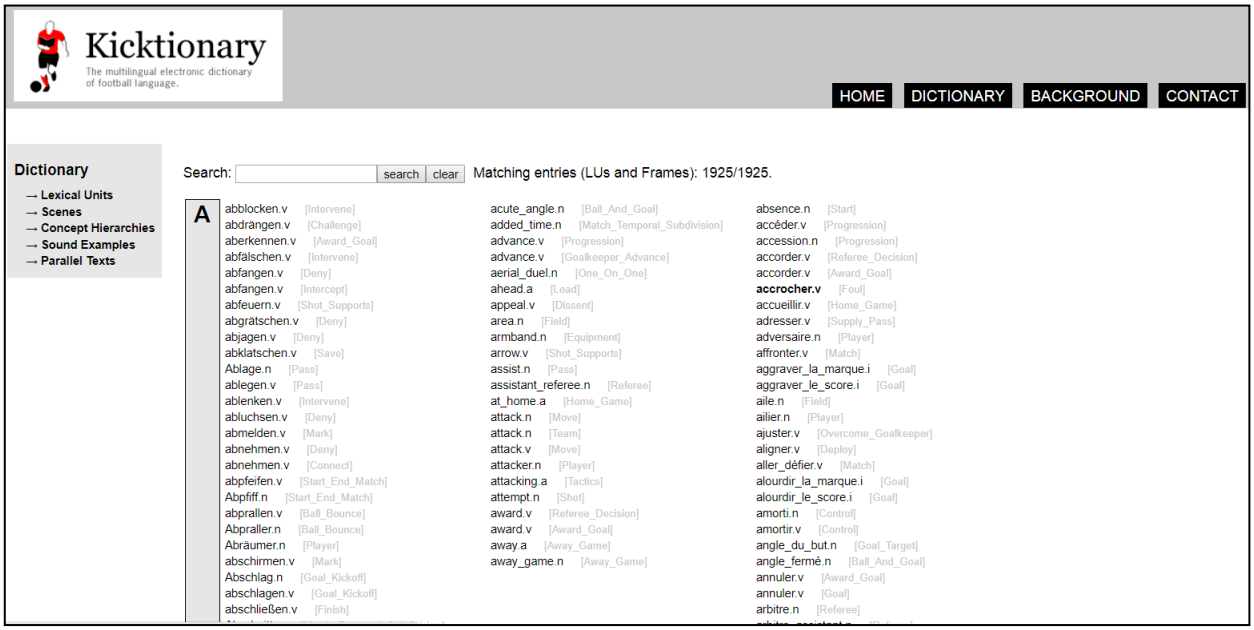


Figura 45: Tela de busca através da UL no Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Na frente de cada UL temos o nome do frame o qual esta UL evoca. Se clicarmos na UL, obteremos a tela de apresentação reproduzida na Figura 46.

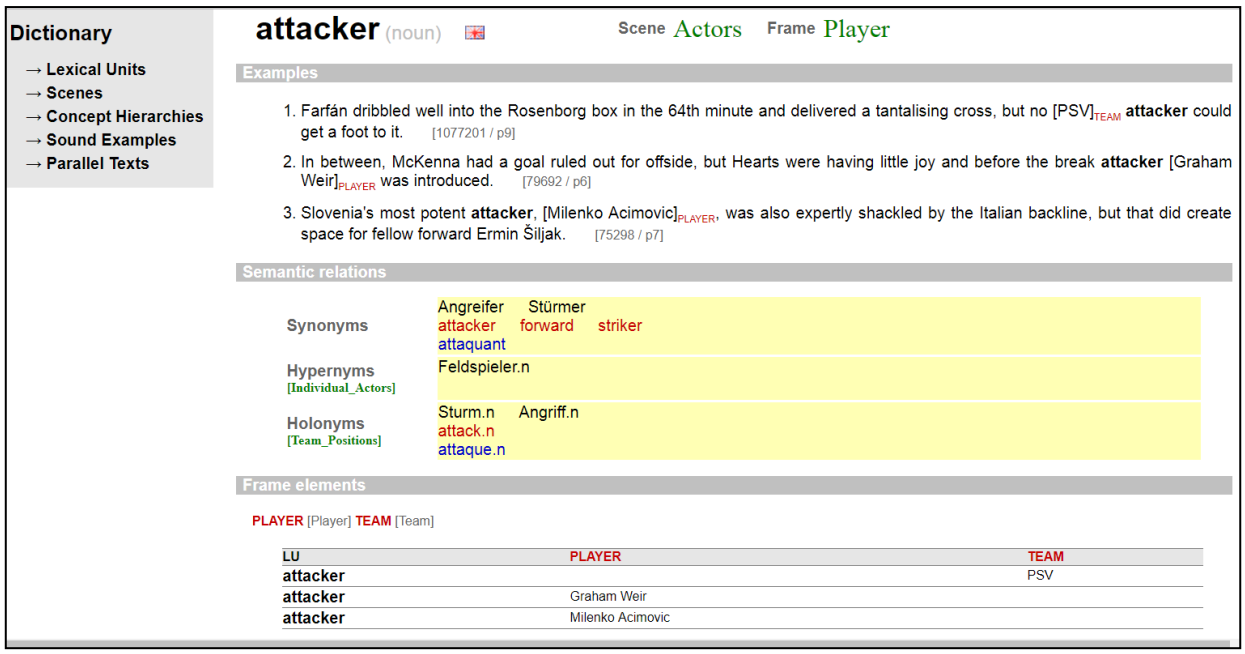


Figura 46: Tela de resultado pela busca da UL “attacker” no Kicktionary.
Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Na Figura 46, temos o nome da UL pesquisada, sua classe gramatical e a língua em que está sendo descrita. Também podemos visualizar a que cena e frame pertence. Logo abaixo, temos os exemplos anotados que instanciam a UL “atacante”, as relações semânticas e os elementos de frame que os compõem.

A segunda possibilidade de busca é por meio das cenas. Ao clicarmos, temos a tela mostrada na Figura 47.

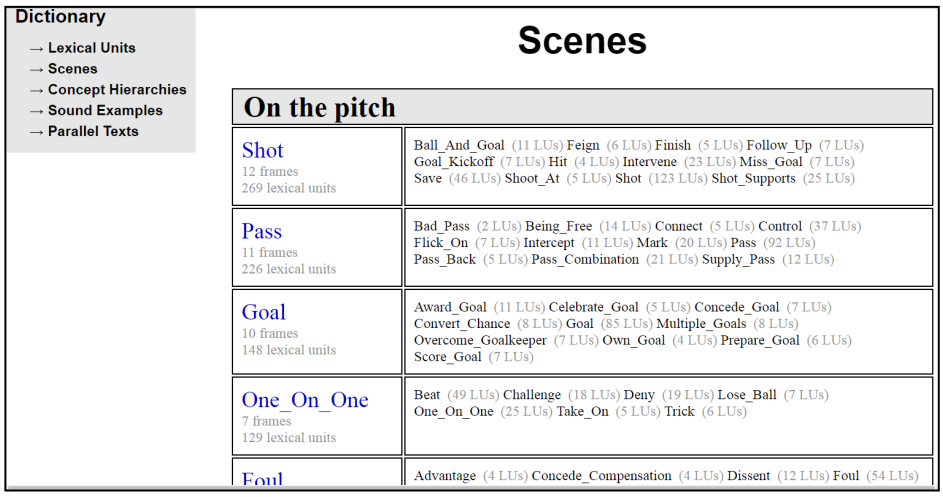


Figura 47: Tela de busca através das cenas no Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Através dessa busca, podemos encontrar as dezesseis cenas e clicar sobre elas, ou sobre os frames que estão agrupados em cada cena, juntamente com as unidades lexicais que evocam esses frames.

A terceira possibilidade de busca é através das relações de hierarquias conceptuais, vide Figura 48.

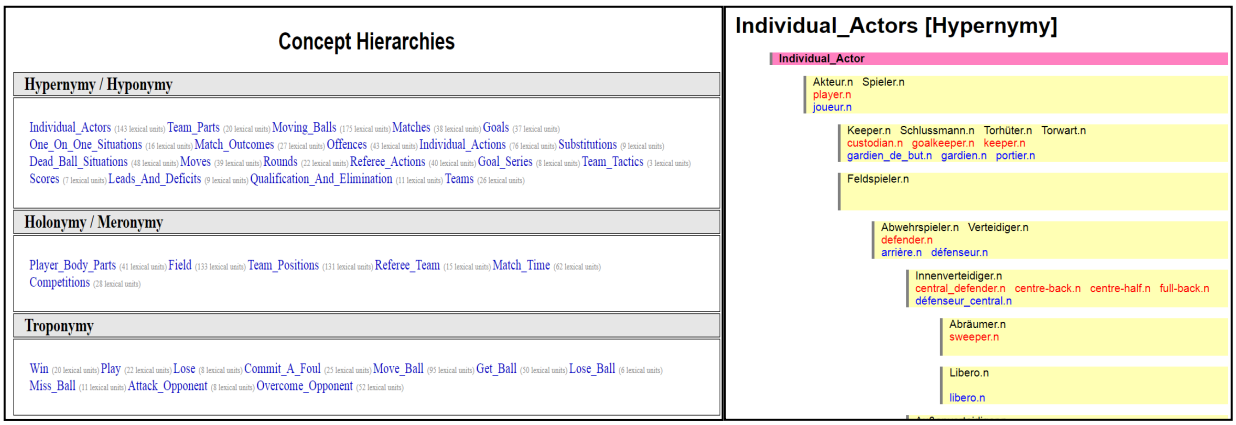


Figura 48: Tela de busca através das relações de hierarquia conceptual no Kicktionary.

Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Ao clicarmos sobre uma das relações, como, por exemplo, na relação Hiperonímia/ Hiponímia, e depois sobre *Individual_Actors*, encontramos uma árvore conceptual descrevendo nas três línguas (alemão, inglês e francês) o nome dos componentes. Podemos acessar cada uma dessas ULs, com suas descrições, seus respectivos frames e cenas.

A quarta possibilidade de busca são os exemplos gravados das narrações, através da qual podemos escutar algumas delas e acessar as ULs, os frames e os arquivos anotados nos quais aparece a UL (vide Figura 49).

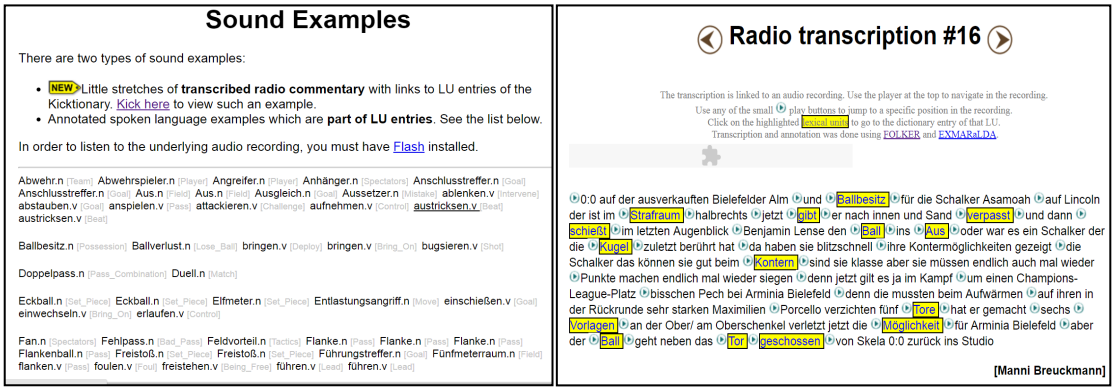


Figura 49: Resultado da busca nos exemplos de fala / narração no Kicktionary.
Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Por último, temos um alinhamento entre os textos, nas três línguas utilizadas no dicionário. Ao clicarmos em qualquer palavra colorida, será exibida a entrada onde aparece esta UL correspondente no Kicktionary e, se passarmos o mouse sobre qualquer uma delas, serão exibidas informações sobre sua estrutura e atribuição de cenário, conforme ilustra a Figura 50.

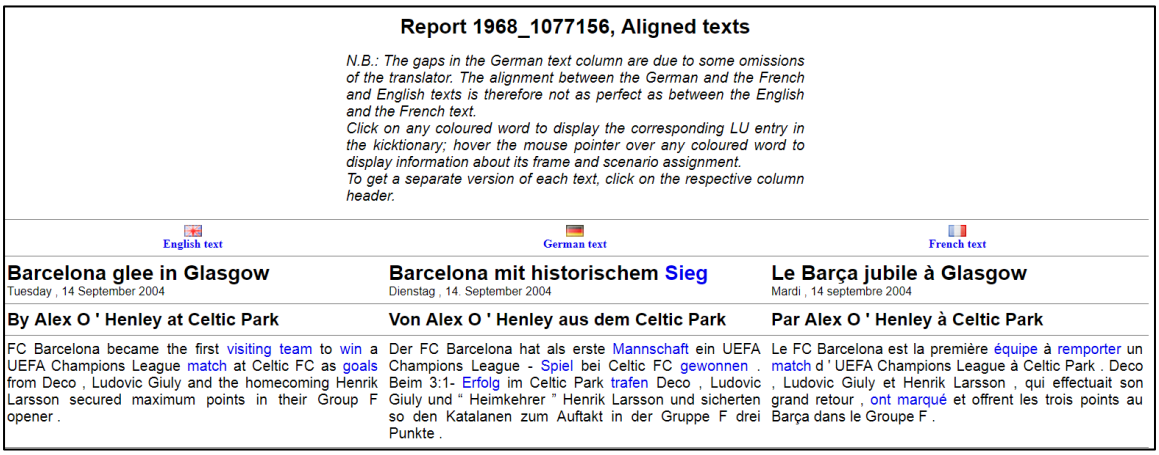


Figura 50: Textos paralelos alinhados com ULs disponíveis para consulta no Kicktionary.
Fonte: http://www.kicktionary.de/index.html. Acesso em 27 de julho de 2017.

Na próxima seção, abordaremos o dicionário *Field*, que também trata do domínio do Futebol.

2.2.5 Field: Dicionário de Expressões do Futebol

O dicionário Field de expressões do futebol é trilíngue (português – inglês – espanhol) e foi organizado com base na noção de cenários. Segundo os dados do dicionário, acessados em julho de 2017, (http://dicionariofield.com.br/), ele apresenta em torno de 900 expressões organizadas em forma de lista de verbetes e de cenários ilustrados para as três línguas. Este produto foi desenvolvido pelo grupo de pesquisa SemanTec do Programa de Pós-Graduação em Linguística Aplicada da Universidade do Rio dos Sinos (Unisinos-RS), coordenado pela professora Rove Chishman.

Partindo do português como língua-fonte, o dicionário futebolístico organiza os contextos de uso, conforme as situações são dadas. Para a produção dessa ferramenta, o grupo de pesquisa baseou-se na Semântica de Frames (FILLMORE, 1982,1985) e, a partir da noção de frame ou cenário, os verbetes foram organizados e estruturados. Apresentaremos as possibilidades de busca e a organização do dicionário, com base nos dados disponíveis no site.

O primeiro passo ao acessar o dicionário no site http://dicionariofield.com.br é escolher uma das três línguas para obter os resultados. Após esse procedimento, o consulente pode escolher como realizar sua busca.



Figura 51: Tela inicial do dicionário de expressões futebolísticas Field.
Fonte: http://dicionariofield.com.br/. Acesso em 28 de julho de 2017.

Conforme vemos na Figura 51, podemos realizar a busca através dos termos futebolísticos, dos cenários ou do campo acima no lado direito da tela que permite escrever o nome do termo para localizá-lo. No primeiro caso, temos a seguinte tela ao acessar (vide Figura 52):



Figura 52: Busca por meio de palavras no dicionário de expressões futebolísticas Field.
Fonte: http://dicionariofield.com.br/. Acesso em 28 de julho de 2017.

Encontramos nela uma lista de termos / expressões futebolísticas de acordo com a língua selecionada. Ao clicarmos em uma delas, aparece a classe gramatical a qual pertence, o áudio da palavra / expressão, assim como sua aplicação em uma sentença, que foi retirada de um corpus, e a tradução para as outras duas línguas. Neste caso, o grupo de pesquisa levou em consideração três aspectos para o processo de equivalência de tradução, podendo ser (i) equivalentes diretos na língua alvo, (ii) equivalentes parciais (quando houver essa situação serão marcados com um asterisco) ou (iii) casos sem equivalentes de tradução (receberá a sigla NT – *no translation*). Se houver um termo variante relacionado ao mesmo significado, este também aparecerá, uma vez que são agrupados na mesma entrada. Há um link no qual o usuário pode acessar o cenário em que se encaixa o termo pesquisado, além de aparecerem do lado direito outras palavras ou até mesmo cenários relacionados, em alguns casos. Quando acessamos o cenário "Participantes", por exemplo, encontramos sua definição, seguida de uma imagem ilustrativa e outros termos relacionados ao mesmo cenário.

A outra possibilidade de busca no dicionário Field é por meio dos cenários, que são caracterizados por representarem eventos, locais dos jogos, ou equipamentos que são utilizados durante uma partida de futebol. No dicionário são definidos 37 cenários que abarcam o domínio do futebol.



Figura 53: Informação encontrada no cenário “Chute” no dicionário de expressões futebolísticas Field. Fonte: http://dicionariofield.com.br/. Acesso em 28 de julho de 2017.

Depois de eleger um cenário, ao clicar nele (Figura 53), encontraremos um áudio do termo, a definição e os seus participantes, e uma imagem ilustrativa. Do lado direito, encontramos outras unidades lexicais que fazem parte do mesmo cenário. Ao clicarmos nelas, aparecerá uma nova tela com os respectivos dados que mencionamos. Trata-se de uma proposta que propõe consultar a linguagem futebolística de forma contextualizada para usuários brasileiros e estrangeiros.

Segundo dados do site, a metodologia empregada para buscar as traduções e coletar os exemplos foi baseada na Linguística de Corpus. A equipe constituiu três corpora comparáveis, com textos extraídos de sites de times de futebol de países como Brasil, Argentina, Espanha, Estados Unidos e Inglaterra. Os recursos computacionais mais utilizados para elaboração do processo foram os softwares AntConc e Sketch Engine.

Na próxima seção, apresentaremos o Dicionário COPA 2014 FrameNet Brasil, que foi desenvolvido pela equipe da Universidade Federal de Juiz de Fora.

2.2.6 Dicionário COPA 2014 FrameNet Brasil

O dicionário Copa 2014 é um recurso computacional lexicográfico, que foi desenvolvido pela equipe da FrameNet Brasil, coordenada pelo professor Dr. Tiago Torrent, em parceria com a equipe da professora Dra. Rove Chishman, da UNISINOS, que desenvolveu o dicionário Field, mencionado e descrito na seção anterior. O COPA 2014 foi configurado como um dicionário eletrônico trilíngue (Português do Brasil, Inglês e Espanhol), para os domínios do Futebol e do Turismo, com o intuito de atender as

demandas de usuários não-especialistas, da impressa esportiva internacional e das pessoas que se envolveram no evento mundial, Copa do Mundo, que foi sediado no Brasil em 2014 (SALOMÃO ET AL. 2011, p. 2). Neste sentido, esta ferramenta foi elaborada com base nas teorias semânticas, nos métodos de pesquisa em corpus e nas tecnologias de hipermídia, relacionando-as.

A FN-Br teve seu início em 2009 com o intuito de expandir a contraparte linguística da rede semântica FrameNet para o português do Brasil. Dessa forma, o banco de dados utilizado para dar prosseguimento a pesquisa no Brasil baseava-se no modelo que foi desenvolvido para a FrameNet do inglês. Com o advento do projeto COPA 2014, foi necessário criar um novo banco de dados, para atender ao domínio específico, distinguindo-se do primeiro por ter sido desenvolvido especificamente para a constituição do dicionário a partir de *corpora* das áreas do Turismo, do Futebol e da Copa do Mundo.

Ao tratar de um domínio específico, o banco de dados da FrameNet do inglês não fornecia frames detalhados que o cobrissem de maneira satisfatória, o que gerou a necessidade de se criarem frames para os domínios do Turismo, Futebol e Copa do Mundo, conforme mencionam Torrent & Ellsworth (2013, p.50):

> (...) os *frames* do banco de dados de domínio específico (DSDB) do Copa 2014 estão sendo criados através de uma abordagem *bottom-up* (ver GAMONAL, 2013, para uma descrição detalhada de tal processo), isto quer dizer que os *frames* foram modelados a partir das sentenças extraídas de *corpora* que cobrem os domínios alvo do dicionário. O resultado de tal abordagem é a criação de *frames* mais detalhados.[48]

Os frames foram criados e modelados para atender às três línguas abordadas no dicionário, tendo em vista que foram considerados como uma representação interlinguística, já que, para os domínios abordados, não havia uma variação significativa entre tais línguas. Este conjunto de frames foi armazenado em um mesmo banco de dados, em três versões que foram traduzidas entre si[49], de modo que, de acordo com a língua escolhida na interface, os resultados eram mostrados para o consulente.

Uma plataforma online foi criada para o dicionário COPA 2014, caracterizando-o como uma ferramenta tecnológica, que pode ser consultado em *smartphones*, *tablets* e computadores que estejam conectados à internet e pode ser acessado no link

---

[48]"(…) frames in the Copa 2014 DSDB are being created via a bottom-up approach (see GAMONAL, 2013 for a detailed description of such a process), this is to say that frames were modeled from the sentences extracted from corpora covering the target domains of the dictionary. The result of such an approach is the creation of more detailed frames."

[49]Para mais detalhes, conferir Peron-Corrêa (2014, p.62).

http://dicionariodacopa.com.br/. Este fator demonstra a superação de algumas restrições e limitações relacionadas aos dicionários impressos.

O produto final está disponível como um aplicativo da web e atende as demandas de usuários não-especialistas, uma vez que a nomenclatura utilizada para definir frames e seus elementos é diferente daquela que é concernente ao campo da Linguística.

Na Figura 54, temos a interface inicial do dicionário, que apresenta a logo e o nome do projeto FrameNet Brasil, seguido das bandeiras correspondentes aos idiomas de acesso, conforme escolha do usuário. Na parte inferior, temos os financiadores do projeto (Universidade Federal de Juiz de Fora, Conselho Nacional de Desenvolvimento Científico e Tecnológico e Fundação de Amparo à Pesquisa do Estado de Minas Gerais).



Figura 54: Interface inicial do dicionário COPA 2014

Fonte: http://dicionariodacopa.com.br

Depois de escolher um dos idiomas de acesso, temos um breve tutorial, com informações sobre a finalidade do dicionário, os domínios que ele abarca e os idiomas de acesso, conforme ilustra a Figura 55.



Figura 55: Tutorial do dicionário COPA 2014

Fonte: http://dicionariodacopa.com.br

Em seguida, temos as possibilidades de busca, oferecidas pelo dicionário. As quatro formas de consultas são: por palavra, por uma frase, por frame (Ver Significado) e, por último, pela rede de relações entre frames (Explorar Rede) (vide Figura 56).



Figura 56: Possibilidades de busca no dicionário COPA 2014

Fonte: http://dicionariodacopa.com.br

A consulta “buscar palavra” segue os critérios que são tradicionalmente adotados por dicionários eletrônicos (cf. PASTOR & ALCINA, 2010), já que oferece uma lista de 1.125 ULs que evocam frames dos domínios do Turismo, Futebol e Copa do Mundo, disponibilizadas em ordem alfabética, de acordo com o idioma selecionado. Ao clicar em uma das ULs, encontramos as seguintes informações: a classe gramatical a qual pertence; a cena que ela evoca, a definição de seu conteúdo e/ou significado e a tradução desta UL para os outros dois idiomas, conforme consta na Figura 57.

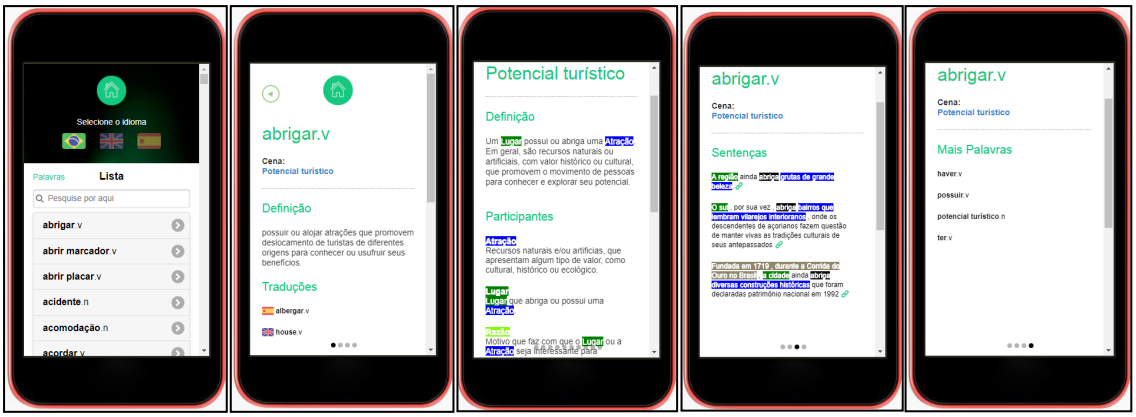


Figura 57: Resultado da consulta por “buscar palavra” no dicionário COPA 2014

Fonte: http://dicionariodacopa.com.br

As bandeiras possibilitam ao consulente alternar entre os léxicos do português, inglês e espanhol. Há várias informações que são disponibilizadas neste modo, por isso, a tela de resultados é dividida em partes (vide Figura 57), as quais podem ser acessadas através de rolagem (*swipe*) lateral. Na segunda parte ou tela, consta o frame (cena) evocado pela UL, neste caso `Potencial_turístico` com sua respectiva definição e os EFs que participam da cena evocada; na terceira tela, temos sentenças anotadas, contendo a UL e a marcação dos EFs em cores para mostrar o uso em dados do corpus; nas últimas telas, há o uso de imagens e/ou vídeos que ilustram o frame evocado. Para que fosse gerada uma lista de equivalentes de tradução, foram anotadas 13.000 sentenças, distribuídas entre as três línguas.

Na segunda opção, temos a busca por “digitar a frase”, vide Figura 58. Essa busca foi planejada como uma estratégia para a resolução de polissemias (cf. CHISHMAN ET AL. 2013). Assim, neste dicionário, se o usuário fornecer um contexto sentencial no qual haja ULs que evocam *frames* constantes da base de dados do dicionário, o algoritmo de desambiguação de *frames* pode inferir o sentido de uma palavra polissêmica, como por exemplo, o verbo “chegar”, que pode evocar os frames de `Chegada` e `Jogar_Mata-mata`, conforme discutido em Peron-Corrêa (2014, p. 43).



Figura 58: Tela de consulta por “digitar a frase” no dicionário COPA 2014
Fonte: http://dicionariodacopa.com.br

No terceiro formato, temos a opção “ver significado”, na qual se encontram a definição de 128 frames trilíngues referentes a Copa, Futebol e Turismo. Ao escolhermos um desses domínios, aparecem, em ordem alfabética, as cenas daquele respectivo domínio. Em cada frame (cena), encontramos a definição e os EFs que aqui são designados de participantes, com suas respectivas definições, segundo temos na Figura 59.

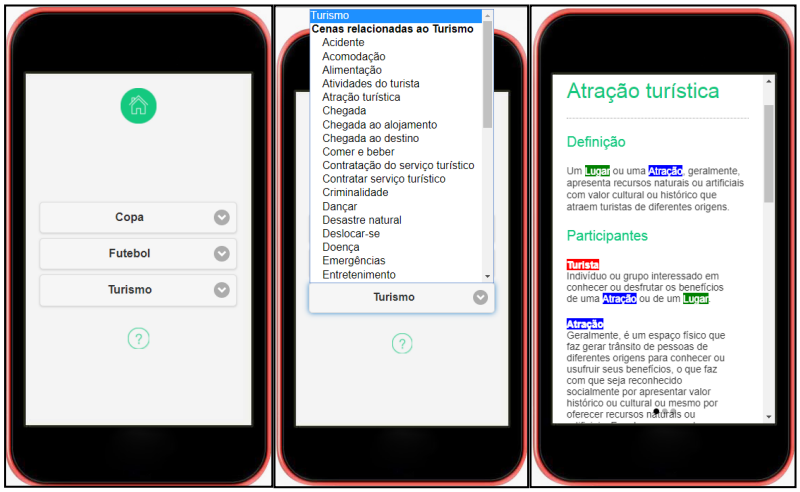


Figura 59: Resultado da consulta por “ver significado” no dicionário COPA 2014.
Fonte: http://dicionariodacopa.com.br

A quarta opção é denominada de “explorar a rede” e mostra as relações que existem entre os frames. A nomenclatura utilizada aqui procura designar de forma mais clara e objetiva para um não-linguista ou não-especialista quais cenas estão interligadas e por meio de qual tipo de relação (vide Figura 60).

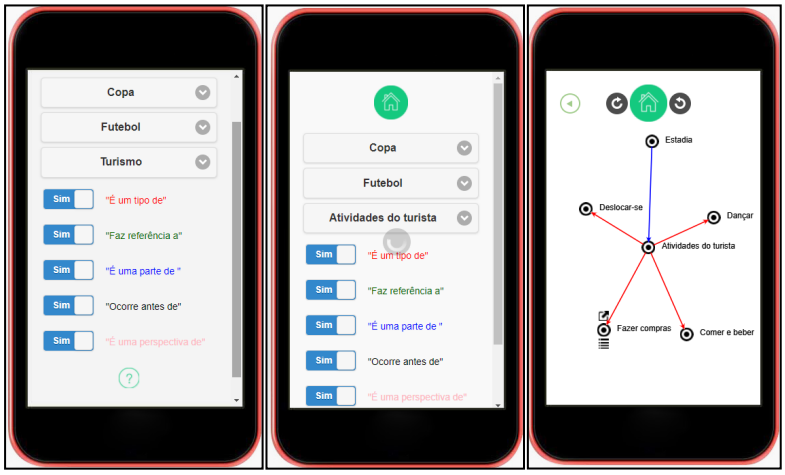


Figura 60: Resultado da consulta por “explorar a rede” no dicionário COPA 2014.

Fonte: http://dicionariodacopa.com.br

Para ilustrar, escolhemos a cena ou frame `Atividades_do_turista`, para visualizarmos quais outros frames estão interconectados por um tipo de relação. As cores das setas auxiliam de acordo com a legenda estabelecida. Na Figura 60, temos uma relação com os frames de `Estadia`, `Deslocar-se`, `Fazer_compras`, `Comer_e_beber` e `Dançar`. Quando clicamos uma vez sobre qualquer um dos frames, temos a opção de reorganizar a rede a partir dele ou de visualizar a tela de resultado do frame, ou seja, aquela que pode ser acessada através da busca *Ver Significado* (vide Figura 61).



Figura 61: Resultado da consulta por “explorar a rede” `Atividades_do_turista` no dicionário COPA 2014.

Fonte: http://dicionariodacopa.com.br

A última opção de acesso, “sobre o projeto”, traz as informações sobre a FrameNet e sua finalidade, os integrantes que participaram do projeto COPA 2014 e as instituições com as quais foram estabelecidas parcerias; os corpora utilizados; e os links que estão relacionados ao projeto.

Na próxima seção, apresentaremos uma tabela comparativa com base nos aspectos lexicográficos expostos por Atkins (2008, p.36), com o intuito de contrastar os critérios lexicográficos que permearam cada uma das ferramentas que foram analisadas nesta seção.

2.2.7 Tabela comparativa dos aspectos lexicográficos das ferramentas descritas

A Tabela 3 resume, para cada um dos recursos analisados neste capítulo, as escolhas que definem o processo lexicográfico proposto por Atkins (2008, p.36) e foi organizada conforme a linearidade presente no diagrama[50], da esquerda para a direita, a qual reflete a cronologia do processo de tomada de decisão. Consideramos que tais critérios foram e são importantes na elaboração de uma ferramenta lexicográfica, já que abarcam cada uma das fases que compõe um todo. Com esta tabela, pretendemos sintetizar o que existe de semelhante e diferente em cada uma das aplicações discutidas, as quais, em alguma medida, compartilham o mesmo aporte teórico da Semântica de Frames.

Após apresentarmos a função Diciopédia do m.knob, desenvolvida nesta tese, novamente apresentaremos a classificação deste dicionário eletrônico segundo os mesmos critérios, de modo a posicioná-lo em relação aos demais recursos baseados em frames.

PRÉ-LEXICOGRAFIA			
e-Dicionários	**Fontes de evidências**		
	Corpus objetivo	*Introspecção Subjetiva*	*Investigação Mista*
EcoLexicon			
DicoEnviro			
DicoInfo			
Kicktionary			
Field			
COPA 2014			

[50] Este diagrama está ilustrado e comentado a partir da página 34 desta tese.

MACROESTRUTURA I – LISTAS DE PALAVRAS			
e-Dicionários	**Tipo**		
	Seletivo	*Abrangente*	
EcoLexicon	■		
DicoEnviro	■		
DicoInfo	■		
Kicktionary	■		
Field	■		
COPA 2014	■		
e-Dicionários	**Lista de Palavras-Núcleo**		
	Não-homográfica	*Parcial. homográfica*	*Totalmente homográfica*
EcoLexicon			■
DicoEnviro		■	
DicoInfo		■	
Kicktionary			■
Field			■
COPA 2014			■
e-Dicionários	**Seleção de Conteúdo**		
	Tipo lexical	*Tipo morfológico*	*Tipo semântico*
EcoLexicon			■
DicoEnviro			■
DicoInfo			■
Kicktionary			■
Field			■
COPA 2014			■
MACROESTRUTURA II – ENTRADAS			
e-Dicionários	**Estrutura da Entrada**		
	Plana	*Em camadas*	
EcoLexicon	■		
DicoEnviro		■	
DicoInfo		■	
Kicktionary	■		
Field	■		
COPA 2014	■		
e-Dicionários	**Status da Entrada**		
	Critério da entrada principal	*Critério da sub-entrada*	
EcoLexicon	■		
DicoEnviro		■	
DicoInfo		■	
Kicktionary	■		
Field	■		
COPA 2014	■		
MICROESTRUTURA I – APRESENTAÇÃO			
e-Dicionários	**Diferenciação**		
	Plana	*Hierárquica*	
EcoLexicon	■		
DicoEnviro		■	
DicoInfo		■	
Kicktionary	■		
Field	■		
COPA 2014	■		
e-Dicionários	**Ordenação**		
	Histórica	*Por frequência*	*Básica*
EcoLexicon			■
DicoEnviro		■	
DicoInfo		■	
Kicktionary			■
Field			■

COPA 2014												
MICROESTRUTURA II – TIPOS DE DADOS												
e-Dicionários	**Interno**											
	Forma						*Conteúdo*					
	Morf.		Orto.		Fon.	Lex.	Explicações			Tipos		
EcoLexicon												
DicoEnviro												
DicoInfo												
Kicktionary												
Field												
COPA 2014												
e-Dicionários	**Externo I**											
	Sintagmático						*Paradigmático*					
	Comp.		Coloc.		Selec.		Gram.		Morf.		Sem.	
EcoLexicon												
DicoEnviro												
DicoInfo												
Kicktionary												
Field												
COPA 2014												
e-Dicionários	**Externo II**											
	Uso						*Relação*					
	Reg.		Atual.		Estilo	Campo	Derivação			Ref. cruzada		
EcoLexicon												
DicoEnviro												
DicoInfo												
Kicktionary												
Field												
COPA 2014												
e-Dicionários	**Outros**											
	Informativo						*Etimológico*					
EcoLexicon												
DicoEnviro												
DicoInfo												
Kicktionary												
Field												
COPA 2014												

Tabela 3: Aspectos lexicográficos das ferramentas descritas

Fonte: Própria autora

Considerando a Tabela 3 sob o ponto de vista do que temos em comum entre as seis ferramentas em análise, podemos pontuar que todas são baseadas em evidências em corpora; são de cunho terminológico, uma vez que trabalham com domínios específicos; além de a seleção de conteúdo ser permeada por dados de tipo semântico, principalmente, devido ao aporte teórico da Semântica de Frames.

Os tipos de dados internos são ancorados principalmente nas bases lexical e morfológica. No que diz respeito ao conteúdo, encaixam-se no que Atkins (2008) denomina de “tipos”, que envolve tanto a linguagem denotativa, quanto conotativa, de significados cognitivos, afetivos, literal e figurado, já que a categoria denominada “explicações” envolve situações como sinônimos, comentários, fórmulas de termos

químicos, referências à listas extratextuais de conjuntos lexicais, como postos militares, pesos, medidas, etc.

Segundo Atkins (2008), os maiores problemas do editor com dados externos se encontram nas relações sintagmáticas, uma vez que não são dados definidos de maneira precisa e exata, mas que se formulam e são empregados pelos falantes no processo de comunicação. Neste sentido, a "preferência selecional" ou as "restrições seletivas" se dão principalmente pela questão do escopo ativo, a partir da seleção que a *headword* pode fazer através de situações de valências e subcategorizações que os falantes nativos realizam sem perceber. Ademais, estudos teóricos têm demonstrado tal pertinência através de corpora e não simplesmente da intuição do falante. De igual modo, as *collocations* estão sendo inseridas nesses dicionários por estarem ancoradas na metodologia da ciência e não simplesmente do fazer de dicionários, já que existem ainda desafios, como, por exemplo, um editor classificar as colocações da língua, assim como elaborar critérios para a atribuição de cada item *multiword* em uma classe apropriada.

Já quanto aos critérios paradigmáticos, os que sobressaem são, principalmente, de cunho semântico, e posteriormente, de cunho gramatical, já que apresentam dados autênticos do uso da língua.

Do ponto de vista do "uso", o fator mais preponderante e que está presente em todos os dicionários é o fator campo, uma vez que tratam de domínios específicos, sem se preocupar com o registro de variantes linguísticas, se é um termo mais obsoleto ou não, se o estilo é poético ou técnico, ou se é usado em outras condições, por exemplo.

Quanto ao aspecto de "relação", o mais plausível para os dicionários analisados trata-se da referência cruzada, que acrescenta e mostra entradas relacionadas, tipos de notas de uso e até mesmo ilustrações pictóricas.

No que tange ao aspecto "informativo" versus "etimológico", consideramos o primeiro mais prevalente, já que dados etimológicos não demonstram relevância nas entradas dos verbetes.

É importante salientarmos, que, segundo Atkins (2008, p.41-42), os aspectos linguísticos têm muito a contribuir para a lexicografia teórica, principalmente, nas decisões sobre a microestrutura dos dicionários, quando, por exemplo, contrapõe-se o "sentido" à "palavra". Assim sendo, quanto mais sistemática a abordagem da lexicografia, mais sistemático será o dicionário finalizado (ATKINS, 2008, p.41). Considerando que a Semântica de Frames é o aporte teórico que estrutura todos os recursos multilíngues eletrônicos analisados, justificam-se, assim, as similaridades apresentadas, devido a

fatores como: considerar o frame como elemento central e norteador das definições, mediante as áreas especializadas; interligar os elementos a outras estruturas, formando, assim, uma rede de conhecimento com base nos princípios semânticos; apresentar definições de *multiwords* e *collocations*, principalmente com base em dados empíricos trazidos pelos exemplos extraídos de corpora; basear as informações, o detalhamento e a definição da UL no contexto evocado, o que facilita no processo de desambiguação, de compreensão de unidades polissêmicas, entre outros.

No primeiro momento, destacamos os critérios em que se aproximavam os dicionários. Também consideramos relevante pontuar os aspectos em que os Dicionários *DicoEnviro* e *DicoInfo* diferem dos demais, como, por exemplo, na lista de palavras nos quais estes são parcialmente homográficos, ou seja, onde um único formulário pode ser compartilhado por várias *headwords,* podendo cada um deles ser polissêmico. Esta escolha inicial acaba interferindo em outras decisões, como na estrutura da entrada dos verbetes ser em camadas, ou seja, onde as sub-entradas podem ser alocadas dentro de uma entrada principal, ou mesmo dentro de outras sub-entradas. Outro critério em que tais dicionários possuem diferença é no quesito de diferenciação hierárquica, o que permite um agrupamento de sentidos de uma forma mais intuitiva satisfazendo: por exemplo, 1, 1a, 1 b, 2, 2a, assim como variantes mais complexas. Além disso, a ordenação é realizada por meio da frequência encontrada no corpus, além de apresentar critérios ortográficos na busca pelas entradas.

Levantados os e-dicionários baseados em Semântica de Frames, redirecionamos o texto, a partir da próxima seção, para a proposição e avaliação da Diciopédia. Nesse sentido, no próximo capítulo, trataremos da metodologia desta pesquisa.

3 METODOLOGIA

Antes de demonstramos os aspectos metodológicos condizentes com esta pesquisa, consideramos importante localizar algumas informações que estão interligadas, de modo sequencial, com o trabalho do qual esta pesquisa faz parte.

Inicialmente, Gamonal (2013) realizou a prospecção de termos e definição de frames para o domínio do Turismo, para o primeiro trabalho lexicográfico aplicado realizado pela FrameNet Brasil, o Dicionário Copa 2014, já apresentado nesta tese. Para o domínio dos Esportes, temos em andamento a pesquisa de doutoramento de Diniz da Costa (2017).

Em ambos os domínios, tivemos a consulta em bibliografia especializada, montagem de corpus e extração de termos a partir da ferramenta *WordSketch.* Nesta etapa, realizou-se a busca por uma determinada palavra, como, por exemplo, “turista” e verificou-se quais verbos mais frequentemente co-ocorriam com ela. Uma vez identificados verbos tipicamente associados aos domínios, realizaram-se novas buscas, desta vez a partir dos verbos, identificando-se os elementos lexicais que mais co-ocorriam com esses, em um processo cíclico. Desse modo, usamos as valências das unidades lexicais para aumentar o próprio repositório terminológico (PERON-CORRÊA, 2014). Posteriormente, consultamos a BabelNet[51] para extrair os equivalentes para o inglês e para o espanhol. Nos resultados obtidos, tivemos um número expressivo para o inglês, enquanto que para o espanhol esse número ficou muito deficitário.

Assim sendo, conforme descrevemos na introdução desta tese, havia uma discrepância muito grande entre o número de termos que a Diciopédia computava em sua versão Alpha para cada uma das línguas (2.316 termos, sendo 1.153 em português, 777 em inglês e 386 em espanhol). Considerando esses aspectos, mas, principalmente, as discussões acerca da forma insuficiente com que os dicionários eletrônicos exploram as potencialidades do suporte digital, tivemos então uma questão motivadora para esta pesquisa com o intuito de desenvolver melhorias para a função Diciopédia do m.knob.

A partir desses dados, surgiu então a necessidade de ampliarmos o repositório multilíngue. Na próxima seção, detalharemos como a continuidade para este processo teve prosseguimento, com enfoque no domínio do Turismo.

---

[51]Esta parte será mais detalhada nas próximas páginas.

3.1 Prospecção de Termos para o Turismo

Para a prospecção de termos para o domínio do Turismo[52], realizamos a constituição de corpora específicos em Português, Inglês e Espanhol. Em relação aos critérios estabelecidos por Atkins (2008), trata-se de **fonte de evidência por meio de corpora objetivos[53]**, uma vez que o aplicativo m.knob é uma ferramenta que visa a atender as três línguas.

Os pré-requisitos estabelecidos por Sardinha (2004) foram seguidos com o intuito de termos um corpus satisfatório, dessa forma, verificamos se o conjunto de dados linguísticos era autêntico e em linguagem natural, originários de falantes nativos e não criados. Ademais, os dados coletados servem ao objeto de estudo e são representativos para o fim ao qual se destinam, além de serem legíveis por computador. Também nos preocupamos em coletar textos nas três línguas que fossem considerados comparáveis, mas não necessariamente paralelos.

Pelo fato de o aplicativo abordar a atividade turística no Brasil, priorizou-se a busca por fontes de dados que tratassem dessa realidade. Os textos foram extraídos especialmente de guias de Turismo, sites de órgãos governamentais de fomento à atividade turística e blogs de viajantes. Primeiramente, era preciso que as informações fossem seguras sobre as atrações, pontos e destinos turísticos, por isso, optamos por fontes de dados oficiais. Em segundo lugar, como o Turismo é uma atividade essencialmente humana, os blogs de viajantes se mostraram relevantes, pois exploram as experiências dos turistas de cada lugar visitado.

A compilação dos *corpora* do Turismo se deu manualmente. A busca por guias disponíveis online, bem como os sites governamentais e blogs de viajantes foi feita na internet, então os textos foram selecionados e copiados em um documento. Sobre o armazenamento, esses dados foram reunidos em arquivos em formato *.txt*, com codificação UTF-8, para que fossem devidamente armazenados na ferramenta gerenciamento e consulta de *corpus* de idiomas, Sketch Engine[54]. Os *corpora* foram

---

[52]Trataremos nesta tese apenas do domínio do Turismo, já que há outros membros do laboratório trabalhando e pesquisando sobre o domínio dos Esportes.
[53]Os dados concernentes à Tabela 3 que dizem respeito ao aplicativo m.knob serão destacados em negrito.
[54]https://www.sketchengine.co.uk/

separados por língua, como podemos ver na Figura 62, constituindo-se um *corpus* para cada língua.

Language	Name	Words
English	m.knob_en	1,121,966
Portuguese	m.knob_pt_br	1,088,885
Spanish	m.knob_es	1,109,667

Figura 62: Corpora m.knob domínio do Turismo

Fonte: https://www.sketchengine.co.uk/

Podemos observar, na Figura 62, que os *corpora* estão balanceados em relação ao número de tokens/palavras, girando em torno de 1.000.000 de palavras em Português, Inglês, Espanhol. A constituição desses *corpora* apresenta as seguintes distribuições: para o Português são 536.918 palavras retiradas de guias turísticos e 551.932 palavras de blogs de viagem; para o Inglês são 589.203 palavras de guias e 502.510 palavras de blogs e para o Espanhol 605.782 palavras de guias e 500.183 palavras de blogs.

Como já havíamos mencionado no capítulo 1, sobre constituição de dicionários de domínios específicos, a ferramenta TERMOSTAT compara o corpus de especialista com um corpus genérico do idioma escolhido e extrai candidatos a termos. Os resultados gerados são disponibilizados em quatro arquivos: (i) *Frequency*: que mostra a lista de candidatos a termos em ordem decrescente de probabilidade; (ii) *Percentage*: que mostra os candidatos a termos agrupados percentualmente de acordo com a suas propriedades morfossintáticas (se são nomes, verbos, compostos de nome + preposição + nome); (iii) *Collocations*: que mostra os candidatos a termos verbais seguidos dos seus argumentos prototípicos no corpus; (iv) *Bigrams*: que mostra os *bigrams* mais frequentes.

Realizamos, inicialmente, a análise dos termos gerados pela frequência conforme apareceram em cada uma das línguas (PERON-CORRÊA E TORRENT, 2017). Para ilustrar, na Tabela 4, consta um pequeno recorte dos dados apresentados em língua portuguesa.

	Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
2	praia	5207	165.02	praia___praias___prainha___praiazinha	Nom
3	rio	3728	123.33	rio___rios___riozinho___riozinhos	Nom
4	hotel	2924	117.59	hotel___hotéis	Nom
5	passeio	2384	113.3	passeio___passeios	Nom
6	pousada	1834	105.41	pousada___pousadas	Nom
7	cidade	4499	103.49	cidade___cidades___cidadezinha___cidadezinha	Nom
8	ônibus	1497	97.71	ônibus	Nom
9	restaurante	1694	93.62	restaurante___restaurantes	Nom
10	dica	1159	85.63	dica___dicas	Nom
11	viagem	1962	84.57	viagem___viagens	Nom
12	prato	1169	78.54	prato___pratos___pratinho	Nom
13	ótimo	921	76.62	ótimo___ótimos___ótimas	Adjectif
14	piscina	1096	75.25	piscina___piscinas___piscininha___piscininhas	Nom
15	bar	1271	74.77	bar___bares___barzinho___barzinhos	Nom
16	parque	1690	71.85	parque___parques	Nom
17	café	1186	71.7	café___cafés___cafezinho	Nom
18	ficar	3949	71.15	___ficarão___ficavam___ficaremos___ficávamos	Verbe
19	pegar	1000	68.76	gar___pegará___pegaram___pegasse___pegarían	Verbe
20	trilha	700	66.6	trilha___trilhas	Nom
21	opção	1340	65.94	opção___opções	Nom
22	endereço	721	64.92	endereço___endereços	Nom
23	atração	649	64.3	atração___atrações	Nom
24	rua	1926	63.24	rua___ruas___ruazinha___ruazinhas	Nom
25	restaurante	789	62.36	restaurante___restaurantes	Adjectif
26	aeroporto	1035	61.77	aeroporto___aeroportos	Nom
27	lindo	691	61.71	__lindas___lindinho___lindíssimo___lindíssimo	Adjectif
28	mapa	836	61.68	mapa___mapas	Nom
29	abraço	668	61.01	abraço___abraços	Nom
30	café de manhã	585	60.97	café da manhã___cafés da manhã	Nom Préposition Nom

Tabela 4: Resultado dos termos gerados pelo TERMOSTAT para o domínio do Turismo em Português.
Fonte: TERMOSTAT (DROUIN, 2003)

De acordo com os dados da Tabela 4, obtivemos cinco (5) colunas especificando as principais informações fornecidas. Na primeira, consta o candidato a termo selecionado; na segunda, aparece a frequência, listada em ordem decrescente; posteriormente, temos os dados sobre a especificação, ou seja, este resultado é o contraste entre a frequência e as propriedades colocacionais que foram encontradas nos corpora de referência com os mesmos parâmetros; em seguida, aparece a lista de variantes ortográficas e, por último, a classe gramatical dos termos selecionados.[55]

Os dados iniciais obtidos para a língua portuguesa foram de 15.217 candidatos a termos, nos quais encontramos substantivos, adjetivos, advérbios e expressões. Para a língua espanhola, encontramos 16.697 candidatos, considerando as diversas classes morfológicas que foram mencionadas anteriormente.

Dentre os aspectos negativos com que nos deparamos com os dados iniciais apresentados, podemos elencar os seguintes: i) termos que constaram mais de uma vez nos resultados, com frequência e especificação distintas; ii) termos em outras línguas, principalmente, nos dados de língua espanhola, com incidência de termos em língua portuguesa; iii) presença com alta frequência de preposições e conjunções.

A partir da tabela inicial dos dados, acrescentamos outras colunas para nos auxiliar na proposta de análise dos termos selecionados pelo TERMOSTAT – Tabela 5 –, de

[55]São os mesmos critérios gerados para as outras línguas.

modo a indicar se o candidato a termo: (a) já constaria de um frame na base m.knob; (b) deveria ser incluído em um frame já constante da base m.knob; (c) evocaria um frame existente na FrameNet de Berkeley que, portanto, deveria ser incluído na base m.knob; (d) demandaria a criação de um novo frame; ou (e) nenhuma das alternativas acima.

Candidat de regroup	Fréquence	Spécificité	ntes orthographi	Matrice	Existe_frame _m.knob	Entraria no m.nob	Frame Berkeley	Novo frame	Observação
música	473	82.82	música___músic	Nom			Performing_arts e fields		
terminal	406	80.17	terminal___term	Nom		transporte			existe o frame,mas a UL não está registrada
comida	470	78.41	comida___comic	Nom		comer_e_beber			existe o frame,mas a UL não está registrada
comer	556	78.2	comer___comes	Verbe		comer_e_beber			existe o frame,mas a UL não está registrada
cascada	306	77.55	cascada___casca	Nom	locais_naturais				
plano	1140	76.91	plano___planos	Nom		planejamento do turista			existe o frame,mas a UL não está registrada
casa	926	76.54	casa___casas___	Nom		acomodación			existe o frame,mas a UL não está registrada
samba	278	76.25	samba	Nom	bailar				
color	451	75.93	color___colores	Nom		stimulus_foco			existe o frame,mas a UL não está registrada
dormir	365	75.66	dormir___dormí	Verbe			sleep ?		
domingo	467	75.37	domingo___dom	Nom		unidade_calêdrica			existe o frame,mas a UL não está registrada
grande	278	75.37	grande	Nom	está em 2 frames: notability e degree				
disfrutar	656	73.63	disfrutar___disfr	Verbe		hacer_turismo			existe o frame,mas a UL não está registrada
dunas	262	73.56	dunas	Nom	locais_naturais				
circuito	396	73.5	circuito___circui	Nom		instalaciones_deportivas			existe o frame,mas a UL não está registrada
blog	256	73.16	blog	Nom			text ?		
paisaje	412	72.39	paisaje___paisaj	Nom		locais_naturais			existe o frame,mas a UL não está registrada
tranquilo	420	71.56	tranquilo___tran	Adjectif				?	

Tabela 5: Dados para análise dos termos selecionados pelo TERMOSTAT

Fonte: própria autora.

Após realizarmos a análise automática do corpus de especialista pela ferramenta TERMOSTAT (DROUIN, 2003), os lexicógrafos da FrameNet Brasil validaram os candidatos a termos, seguindo os critérios abaixo, que compõem um processo de três estágios:

1. Procura-se o candidato a termo na base corrente do m.knob:
    a. Caso ele já tenha sido incluído, ele é considerado válido.
    b. Caso contrário, passa-se ao estágio 2.
2. Averiguam-se no corpus sentenças que contenham o candidato a termo e realiza-se uma pré-anotação semântica das sentenças:
    a. Caso as anotações se enquadrem em um frame já incluído na base corrente do m.knob, o termo é considerado válido.
    b. Caso contrário, passa-se ao estágio 3.
3. Discutem-se com a equipe de lexicógrafos as anotações não-conformes:
    a. Caso elas constituam um frame relevante, o termo é considerado válido.
    b. Caso contrário, ele é descartado.

O processo inicial para a definição dos termos começa pela extração em corpus de verbos que são frequentemente associados aos termos nominais prototípicos como “turista”, “atração”, “museu”, tomando-os como sujeito ou objetos. Assim, esses verbos são pré-anotados e os frames são criados. Na sequência esses próprios verbos serão utilizados para encontrarmos no corpus novos termos nominais, adjetivais e assim, por

conseguinte, formando o processo *bottom-up* (do termo para o frame*).* Quando temos o frame, o processo desenvolvido é o *top-down,* no qual, a partir do frame, realizamos busca de outras palavras no corpus que a intuição do linguista leva a crer que seja um termo que evoque tal frame.[56]

Essa metodologia proposta nos auxiliou na identificação de termos do domínio especializado, permitindo-nos compreender e identificar a diferença do uso de tais palavras em domínio mais genérico ou específico, visto que a constituição, seleção e validação de tais termos requerem maior cuidado, já que não se trata apenas de um processo de listagem, e, sim, de compreensão do processo que envolve os empréstimos linguísticos dentro de um contexto.

3.2 Análise de Interação Usuário-Interface

A fim de verificar se os parâmetros metodológicos propostos na Diciopédia são relevantes e satisfatórios para o consulente, no âmbito de sua apresentação em plataforma *mobile*, aplicamos um questionário (ANEXO 1), elaborado por meio da ferramenta Formulários do Google e realizamos também testes de usabilidade, através da aplicação Appsee para 38 (trinta e oito) estudantes do curso de Letras da Universidade Federal de Juiz de Fora, que na grande maioria, foram alunos que estavam na primeira semana de aula referente ao seu primeiro período, ou seja, tinham acabado de entrar para o curso de Letras. Ressaltamos também que se mostraram disponíveis para participar dessa pesquisa. Ao serem convidados, tanto por correio eletrônico como pessoalmente, um dos critérios frisados foi se tal participante tinha familiaridade com uso de aplicativos de celulares. Os testes e o questionário foram aplicados no Laboratório FrameNet Brasil. Os dados obtidos foram analisados para discutir em que medida as informações apresentadas foram percebidas pelo usuário.

Para averiguarmos a interação do usuário com o aplicativo Diciopédia, utilizamos uma plataforma de análise de aplicativos denominada Appsee[57], que fornece dados detalhados do comportamento dos usuários que utilizam aplicativos, como a Diciopedia, conectados à plataforma de avaliação. Para promover essa conexão, o envio,

---

[56] Para maiores informações conferir PERON-CORRÊA (2014) e GAMONAL (2013).
[57] https://www.appsee.com

armazenagem e análise das interações do usuário com o aplicativo, insere-se um SDK (*Software Development Kit*) na linha de código do aplicativo que se deseja avaliar. Uma vez definido o escopo avaliativo que interessava a esta tese, o SDK fornecido pela plataforma Appsee foi inserido, pelos integrantes da equipe de desenvolvimento do Laboratório FrameNet Brasil, em versão de teste fechado do m.knob na plataforma Android.

Conforme as informações disponíveis em seu site, o Appsee permite que os editores e desenvolvedores de aplicativos para dispositivos móveis rastreiem, compreendam e melhorem a experiência do usuário em seus aplicativos e atuem sobre os *insights* fornecidos pela plataforma.

Dentre esses dados, temos as gravações das interações dos usuários, o que permite visualizar a jornada dos consulentes no aplicativo, já que cada toque, deslize e ação de cada tela é gravada, possibilitando constatar os *insights* instantâneos da experiência do usuário no aplicativo. Além disso, ele compila um mapa de toque para cada tela do aplicativo, denominado *touch heat map*, permitindo, assim, detectar onde os usuários estão concentrando sua atenção; verificar os gestos que não respondem em cada tela, assim como os elementos intuitivos de interface do usuário.

Outro dado disponível é o recurso de fluxos de usuários, que fornece uma visão geral de como os usuários estão navegando no aplicativo. Dessa forma, podemos detalhar uma tela específica para descobrir estatísticas importantes, como o número de sessões, impressões e as telas de entrada ou saída correspondentes. Esses elementos nos auxiliaram na avaliação da interface do usuário.

Todos os 38 voluntários tiveram os dados de interação de suas sessões armazenadas pelo Appsee. A partir desse quantitativo, a plataforma faz uma amostragem randomizada de 30 sessões de acesso e aplica o percentual de redirecionamento gerado pelo fluxo de usuários para a geração dos *touch heat maps*. Assim, os seguintes quantitativos de sessões foram usados para cada tela acessada: 30 para a tela de abertura do m.knob, 20 para a tela de chegada da Diciopédia e do verbete e 16 para uma nova busca na Diciopédia. Apesar de tal amostragem reduzir o número total de sessões no cômputo dos *touch heat maps* e demais relatórios de usabilidade, tal redução não compromete, em absoluto, a validade do teste de usabilidade. A bibliografia na área define como 5 participantes o número adequado para testes genéricos de usabilidade. Tal número deve crescer até um máximo de 20 usuários, quantitativo capaz de revelar com o maior grau de detalhamento possível, o comportamento dos usuários

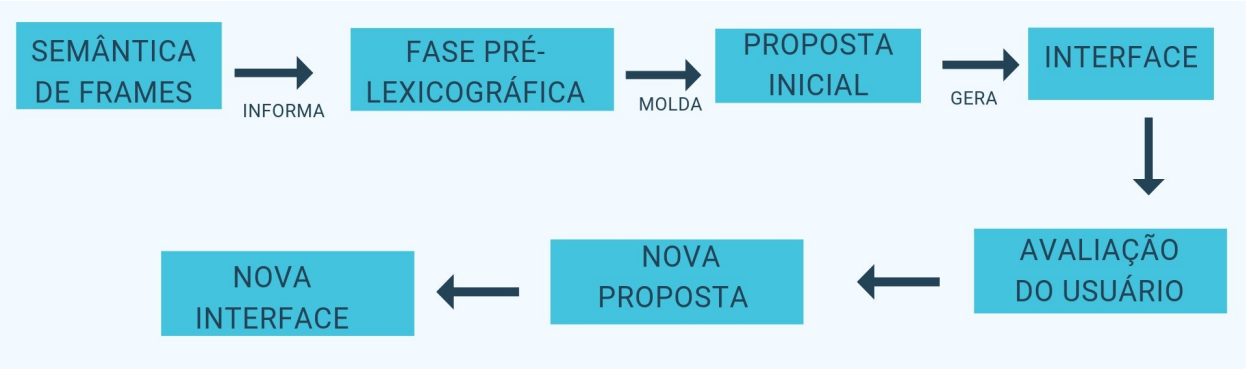
em sua interação com a interface (vide ALROOBAEA E MAYHEW, 2014 para uma discussão e revisão da literatura pertinente à definição de número de testadores nos experimentos de usabilidade).

Nenhum dado pessoal do usuário é armazenado pela plataforma e todas as interações com o aplicativo foram feitos a partir de um mesmo dispositivo móvel, telefone celular, fornecido pela pesquisadora. Assim, não houve qualquer risco de exposição de dados dos voluntários.

3.3 Retomada e esquematização do processo lexicográfico

Antes de finalizarmos o capítulo de metodologia, retomaremos os principais aspectos desta tese, tais como, o **problema de pesquisa**, que foi formulado através da seguinte pergunta: Quais parâmetros metodológicos devem guiar o desenvolvimento de dicionários eletrônicos temáticos multilíngues baseados em recursos lexicais? Sendo que a **hipótese** proposta está relacionada com a relevância da estrutura de dados apresentada ao usuário, já que se correlaciona diretamente com as possibilidades ofertadas pela teoria linguística que dá sustentação ao recurso lexicográfico, considerado o seu propósito. Em relação ao **objetivo principal** desta tese, pretendemos fornecer uma resposta quantificável do que é útil e relevante para o usuário de dicionário eletrônico temático multilíngue baseado em frames, a partir da testagem de metodologia desenvolvida para a elaboração de tais dicionários.

Considerando esses três aspectos importantes da tese, podemos propor o seguinte esquema que resume o percurso teórico e lexicográfico concernente à ferramenta Diciopédia, do aplicativo m.knob.



Conforme ilustra o esquema, a Semântica de Frames (teoria norteadora do dicionário) informa quais serão as principais escolhas a serem desenvolvidas na fase pré-lexicográfica. A partir de tais dados, foi moldada uma proposta inicial, gerando também uma interface de acesso para o usuário. Propusemos uma avaliação de usuários não-especialistas, com base na qual – vide os dados discutidos no capítulo 5 desta tese – foi possível propor uma nova interface – conforme consta no capítulo 6 – que procurasse atender as demandas do usuário.

No próximo capítulo, apresentamos e detalhamos as informações referentes à função Diciopédia.

4 DESENHO DA FUNÇÃO DICIOPÉDIA DO M.KNOB

Neste capítulo, descrevemos brevemente o aplicativo m.knob, para, em seguida, apresentar a proposta, desenvolvida nesta tese, para a função Diciopédia, dicionário eletrônico temático multilíngue, caracterizado por ser um repositório de termos relevantes dos domínios do Turismo e dos Esportes, sendo, portanto, do **tipo seletivo** no que tange aos critérios de Macroestrutura I estabelecidos por Atkins (2008), assim como os demais recursos que foram apresentados no capítulo 2. Cada um dos itens desenvolvidos será descrito com base nos critérios teóricos que foram expostos nos capítulos iniciais. Esse aplicativo apresenta outras funções, como o guia turístico (sistema de recomendação) e o tradutor (em desenvolvimento)[58].

Em linhas gerais, o sistema de recomendação baseado em busca semântica é um assistente de viagens, cujo objetivo principal é recomendar atrações turísticas e diversos locais como hotéis e restaurantes. Ele se apresenta na forma de um chatbot, ou seja, o turista interage com o aplicativo a partir de digitação de texto numa interface conversacional. O chatbot aparece na figura do GregBot, um avatar presente na interface inicial do aplicativo.

A partir das recomendações, o usuário poderá acessar mais detalhes sobre os dados sugeridos, como mapas, horários, site oficial, descrição, assim como avaliar o nível da recomendação. Essas informações são extraídas automaticamente de base de dados abertos disponíveis online, como Google Places e WikiData; já o processamento semântico, o armazenamento e o tratamento dessas informações são realizados pela FrameNet Brasil, através de um léxico multilíngue fundado em frames (vide DINIZ DA COSTA ET AL. 2018 para uma apresentação detalhada do sistema de recomendação).

A outra função, o tradutor híbrido (intérprete pessoal), ainda está em fase inicial de desenvolvimento. A proposta inicial parte da premissa de que a FrameNet pode contribuir significativamente para a Tradução Automática (TA), pois, em suas bases de dados, encontram-se o conhecimento acerca dos *frames* e da gramática de uma língua (PERON-CORRÊA et al, 2018).

A tela inicial do aplicativo mostra o GregBot, iniciando uma conversa com o intuito de direcionar o consulente em sua pesquisa, conforme mostra a Figura 63. Assim sendo,

---

[58]Essas funções estão sendo desenvolvidas por outros alunos de pós-graduação e, por isso, não serão detalhadas nesta tese.

para consultar a Diciopédia, temos duas formas possíveis: (i) clicar na parte superior do lado direito da tela no ícone em que consta “AZ”; ou (ii) clicar direto no botão em que consta o nome “Diciopédia”, localizado na parte inferior da tela, ao lado das outras duas funções do aplicativo “descobrir lugares” e “tradutor”. Na sequência, temos a tela seguinte na qual aparece a instrução “tente digitar algo!”, onde o usuário pode escrever uma palavra ou expressão dos domínios atendidos para chegar ao resultado esperado.

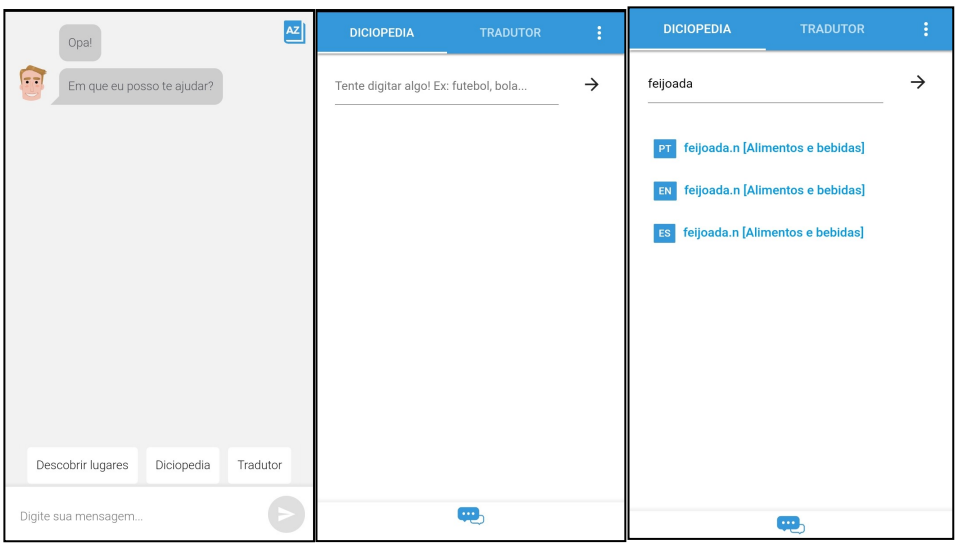


Figura 63: Tela inicial do aplicativo
Fonte: http://www.mknob.com/

Para ilustrar, digitamos a UL “feijoada”, e, inicialmente, obtivemos o resultado de busca nas três línguas (tal informação está presente no ícone que vem antes das palavras, com os códigos ISO 639-1 para cada língua). Ao clicarmos em um dos resultados teremos uma nova tela com as informações relativas à UL pesquisada. Desse modo, ressaltamos, assim, que a **estrutura da entrada** referente à macroestrutura II, é caracterizada como **plana**, ou seja, cada item tem a sua própria entrada principal, assim como a dos recursos EcoLexicon, Kicktionary, Field e Copa 2014. A informação presente entre colchetes é o nome do frame que cada UL evoca, ou seja, neste exemplo, a UL “feijoada.n” evoca o frame de `Alimentos_e_bebidas`, já que cada card apresentado no m.knob ilustra uma UL que, por ser uma UL, está vinculada a um frame.

Nas próximas seções, mostraremos de maneira mais detalhada a estruturação do verbete e das demais informações que compõem os resultados obtidos na Diciopedia.

4.1 Estruturação do Verbete

Como nossas entradas são definidas por unidades lexicais, ou seja, o emparelhamento de forma e significado, os frames que são evocados por tais unidades lexicais auxiliam no processo de desambiguação de homógrafos, conforme podemos verificar na consulta do termo “check-in”, vide Figura 64. Neste caso, o termo “check-in.n” evoca dois frames diferentes: `Chegada_ao_alojamento` e `Partida_do_turista_localidade`. Através da configuração das informações primordiais por meio de redes semânticas, os sentidos podem ser localizados por meio de frames, que podem inclusive especificar e até mesmo auxiliar na delimitação de cada unidade lexical.

A estrutura geral do verbete apresenta na entrada o lexema, precedido pela língua e seguido de sua categoria gramatical (*n* se refere ao nome, *v* ao verbo, etc.), assim como o nome do frame que esta unidade lexical evoca, conforme mostra a Figura 64.

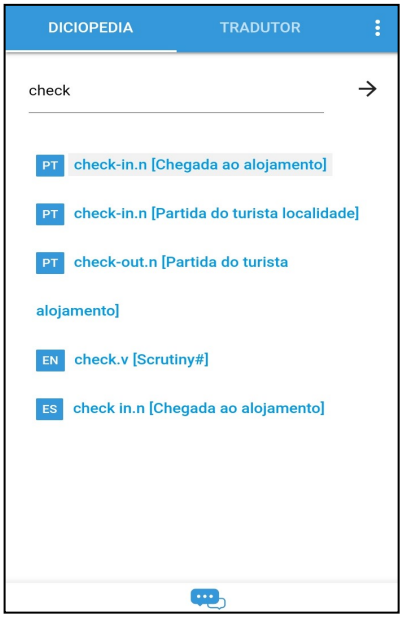


Figura 64: Resultado de busca do termo “check-in”
Fonte: http://mknob.com/. Acesso em 20/12/2018.

Considerando a tipologia de macroestrutura proposta por Atkins (2008), **a lista de palavras da Diciopédia é do tipo totalmente homográfico**, como os dicionários EcoLexicon, Kicktionary, Field e Copa 2014. As ULs e suas definições são apresentadas a partir de sua relação com o frame, caracterizando-se, portanto, como sendo a **seleção de conteúdo de tipo semântico** como dos demais recursos que foram analisados.

Retomamos, também neste sentido, a proposta de KILGARRIFF (1997), sobre a importância de as palavras serem agrupadas de acordo com seus aspectos semânticos, assim como tendo como base a convergência estatística de seus usos variados que aparecem no corpus.

Após os resultados obtidos, ao clicarmos na UL desejada, encontramos sua definição baseada no modelo enciclopédico – Figura 65 –, sendo o **status da entrada** caracterizado como **critério da entrada principal**, como nos recursos EcoLexicon, Kicktionary, Field e Copa 2014. Quanto à microestrutura I, no que tange à **diferenciação**, temos a **estrutura plana**, no qual todos os sentidos terão status igual, assim como os dicionários mencionados anteriormente. A **ordenação**, também presente na microestrutura I, se enquadra no **critério básico**, uma vez que não é por frequência e nem por aspecto histórico.

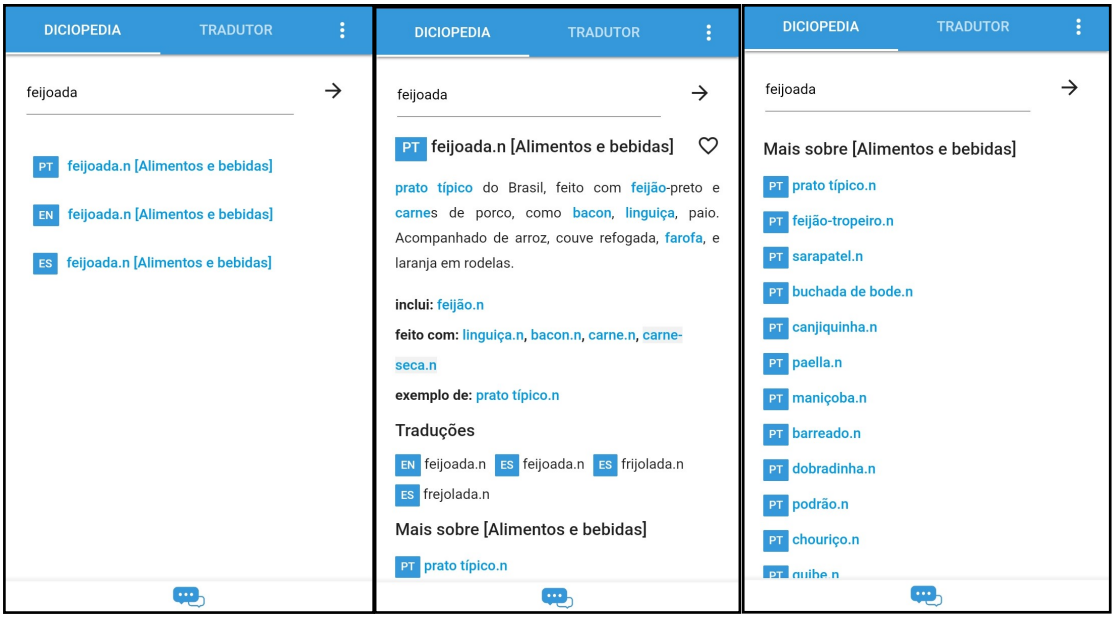


Figura 65: Resultado de busca da UL “feijoada”
Fonte: http://mknob.com/. Acesso em 20/12/2018

Outro dado que aparece na sequência, conforme mostra a Figura 65, são as estruturas qualia ternárias que complementam as informações com uma descrição semântica, como, por exemplo, “inclui”, “feito com” e “exemplo de”[59]. Os nomes das

[59]A descrição e o detalhamento de todas as estruturas qualia desenvolvidas e utilizadas na FrameNet Brasil estão sendo descritas na tese de Diniz da Costa (em andamento).

relações qualia registrados no aplicativo são intuitivos de modo que atendam ao público não especialista.

Com base nesses dados e nas buscas da Diciopédia, os **tipos de dados internos** concernentes à microestrutura II, são de cunho, principalmente, **lexical e morfológico no que tange à forma**. **Morfologicamente** foram incluídas, além das formas de verbete, as formas flexionadas das ULs, as quais não são mostradas na tela, mas são armazenadas no banco de dados. Quanto ao quesito **conteúdo**, se classifica como **apresentação de tipos**, que pode incluir significado literal, figurado, denotativo, conotativo, cognitivo, afetivo e assim por diante, como todos os demais recursos analisados.

Considerando os tipos de dados externos I, sob o viés **sintagmático**, a Diciopédia apresenta significados de *collocations, como, por exemplo, "jardim botânico", "jardim zoológico",* que atendem às **restrições seletivas**, a partir de questões como as estruturas qualia (exemplificadas anteriormente) e da estrutura argumental sintática preenchida pelos complementos. Isso porque, para a definição das ULs que integrarão os frames do domínio m.knob, análises prévias das ocorrências dessas ULs em corpora têm papel importante na definição de seu sentido. Assim, a UL verbal "atrair" consta da Diciopédia em seu sentido para o Turismo, constante de sentenças como: "*As belas praias cariocas atraem turistas do mundo inteiro*". Novamente, ainda que essas informações não sejam apresentadas na tela do verbete, elas constituem o trabalho lexicográfico que dá sustentação ao recurso.

No aspecto **paradigmático**, a Diciopédia traz informações, principalmente, que tangenciam os critérios **semânticos**, uma vez que nos ancoramos teoricamente nos frames, e também alguns **gramaticais**, como a classe gramatical do termo pesquisado, conforme podemos verificar na Figura 66.

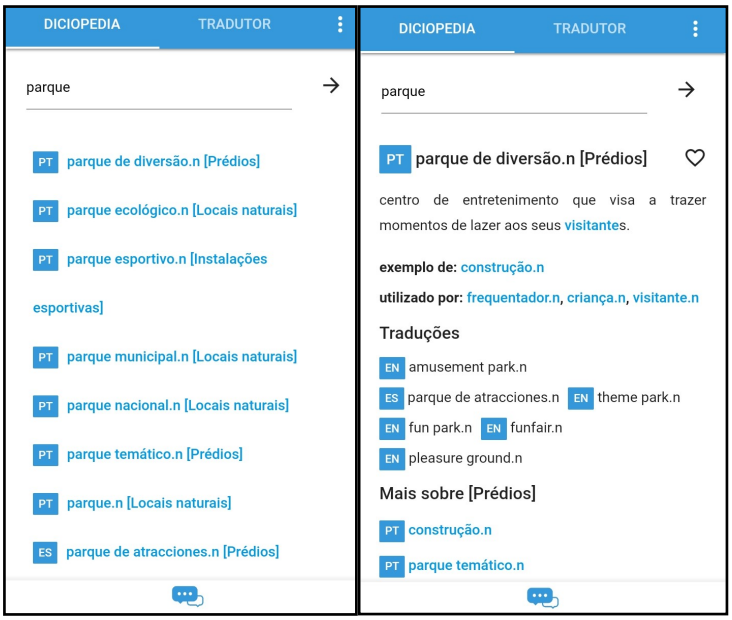


Figura 66: Resultado de busca da UL “parque” no mknob.
Fonte: www.mknob.com. Acesso em 09/01/2019

Procuramos também inserir palavras sinônimas, com o intuito de possibilitar um maior leque de buscas dentro de um mesmo contexto. Através da configuração das informações primordiais por meio de redes semânticas, devido ao aporte teórico da Semântica de Frames, os sentidos podem ser localizados por meio de frames, que podem inclusive especificar e até mesmo auxiliar na delimitação das especificidades de cada unidade lexical. Como exemplo, temos a *multiword* “parque de diversão” e que tem como sinônimo disponível na busca “parque temático”.

Mesmo considerando os locais, e nome de atletas importantes para os domínios do Turismo e dos Esportes, optamos por não inserir nomes próprios. Entretanto, palavras que remetem a pratos típicos, mesmo que sendo típicos apenas do Brasil, foram introduzidos, conforme ilustramos na imagem do card sobre “feijoada.n”.

Quanto ao **uso**, podemos classificá-lo como **preenchendo dados sobre um campo específico**, assim como os demais recursos analisados, já que estamos tratando do domínio do Turismo e dos Esportes.

Outra informação subsequente do verbete são as traduções[60] para inglês e espanhol de termo ou expressões pesquisadas, caso a consulta seja em português. As

[60] Considerando ser uma discussão longa e muito importante sobre a definição do termo “equivalentes” de tradução, conforme discutimos na dissertação de mestrado (PERON-CORRÊA,2014), elegemos o termo “tradução” para mostrar os significados para as outras línguas, já que seria um termo menos especializado e também mais direto, até mesmo para não-linguistas.

traduções que são apresentadas na Diciopédia são geradas automaticamente, seguindo duas metodologias distintas: (i) para nomes de entidades, ou seja, que se referem às pessoas, objetos e lugares, os dados são extraídos automaticamente de uma base de dados ligados, denominada BabelNet e à posteriori são validos pelos linguistas do projeto FrameNet Brasil. A BabelNet é um recurso computacional de dados lexicais que liga informações extraídas da Wikipédia à WordNet e à FrameNet. Conforme descreve Navigli (2012, p.2), a BabelNet tem como resultado *"um 'dicionário enciclopédico' que fornece babel synsets, ou seja, conceitos e entidades nomeadas lexicalizados em muitas línguas e conectados através de uma grande quantidade de relações semânticas."* Este recurso visa fornecer uma cobertura lexicográfica e enciclopédica ampla, incluindo 14 milhões de entradas em 271 línguas. A segunda metodologia (ii) está relacionada aos verbos e aos nomes que indicam evento: os dados são calculados tomando como base as anotações lexicográficas do projeto FrameNet Brasil, que consideram as valências sintáticas e semânticas, de acordo com o uso destas em textos reais (PERON –CORRÊA ET AL., 2016). Novamente, parte da estrutura de dados da FN-Br, que não é diretamente mostrada no verbete, impacta a configuração deste.

Outras informações que aparecem para os consulentes como parte da pesquisa são as palavras relacionadas, ou seja, que estão presentes no mesmo frame, conforme temos na última tela da Figura 66. A ordem dessa listagem está baseada no número de qualia ternários em comum que elas apresentam, ou seja, quanto mais relações qualia compartilhadas entre a UL da entrada e a UL da lista com uma terceira UL, maior será o ranqueamento dessa UL em relação às demais. Todas as palavras de um mesmo frame são listadas nessa seção do verbete.

Esses dados adicionais, assim como a própria estrutura qualia inserida na complementação dessas informações podem ser considerados como sendo de cunho **informativo**, e não etimológico. Há também, a **relação de referência cruzada**, em que através das ferramentas de hipertexto (hiperlinks), temos uma referência cruzada, como veremos na próxima seção.

4.2 Ferramentas de Hipertexto

Ainda na função "Diciopédia", um dos aspectos relacionados à função da microestrutura II, concernente ao dado externo II, estabelecido por Atkins (2008, p.36), é a **referência cruzada**, conforme foi mencionado anteriormente. No aplicativo

desenvolvido, tais referências aparecem como hiperlinks, em cor azul (Figura 67), possibilitando ao usuário uma navegação intuitiva nessa rede de significados.

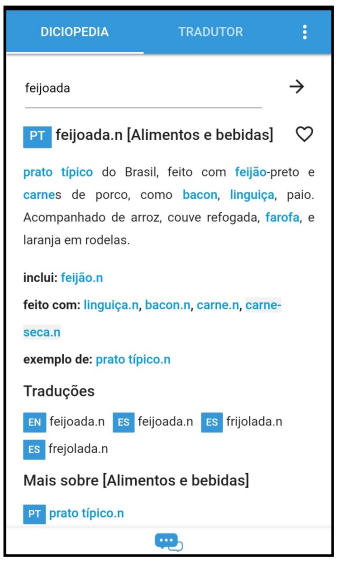


Figura 67: Resultado de busca da UL “feijoada.n”
Fonte: http://mknob.com/. Acesso em 27/12/2018

A partir dessa tela (Figura 67) consultamos alguns dos termos que estão em azul, e obtivemos os seguintes resultados respectivamente, ao clicar em, por exemplo, “prato típico”, “feijão”, “carne”, conforme mostra a Figura 68.

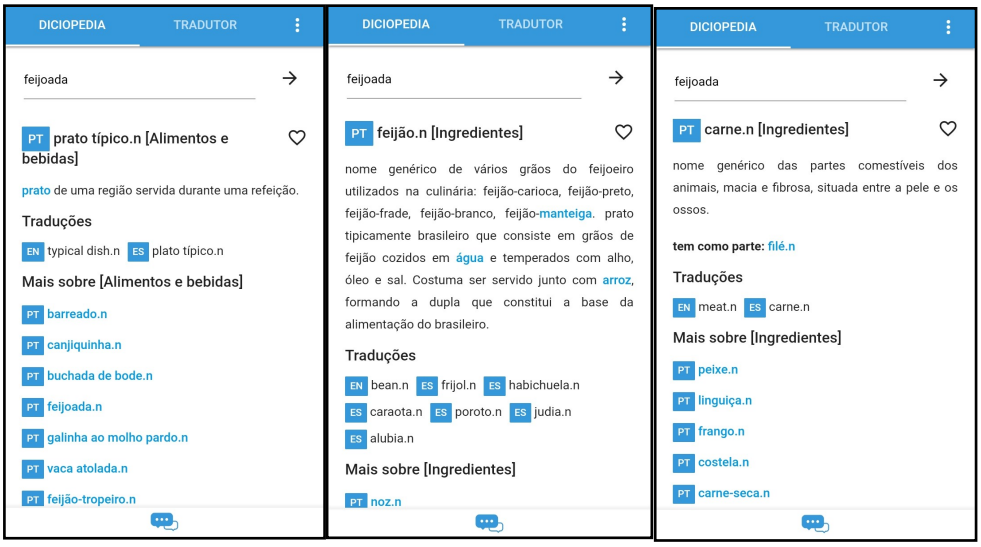


Figura 68: Resultado de acesso às ferramentas de hipertexto.
Fonte: http://mknob.com/. Acesso em 20/12/2018

Cada um dos hiperlinks disponíveis em cada consulta é clicável e, ao ser acessado, mostrará os resultados de definição, de estrutura qualia disponível, das traduções, e de outras palavras relacionadas ao termo consultado, ou seja, novas consultadas podem

ser realizadas a partir de palavras, expressões e termos disponíveis em uma única busca. Esse resultado pode levar até outras buscas e resultados a partir de dados interligados pela rede de frames e estruturas qualia ternárias modeladas na base de dados da FrameNet Brasil.

4.3 Ferramentas Colaborativas

Além dos itens e ferramentas que já foram descritos anteriormente, temos também na Diciopédia algumas ferramentas colaborativas, uma vez que o usuário pode interagir com o aplicativo.

Ao tentar buscar uma palavra ou termo que não conste no banco de dados e seja pertinente aos domínios do aplicativo, o usuário pode sugerir que este termo seja incluído, como mostra a Figura 69, e a equipe da FN-Br irá avaliar e validar o termo em questão.

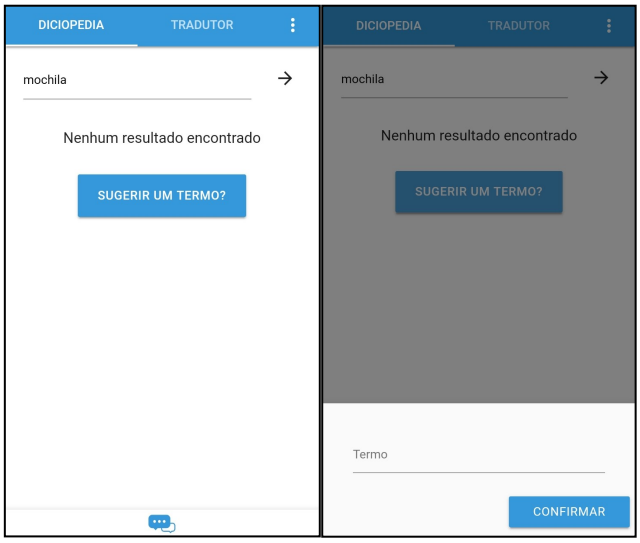


Figura 69: Tela de termo não encontrado

Fonte: http://mknob.com/. Acesso em 27/12/2018

Por exemplo, ao digitarmos o termo “mochila”, não obtivemos nenhum resultado encontrado na Diciopédia. A partir desse dado, o usuário pode sugerir o termo. Caso ele faça essa opção, na parte inferior do aplicativo aparecerá a opção “Termo”, que deverá ser digitada e em seguida, basta o consulente clicar em “confirmar”.

É importante ressaltarmos que cada uma das informações descritas da função Diciopedia, faz parte do que Atkins (2008) denomina de tomada de decisões lexicográficas, uma vez que cada um dos itens que aparece e até mesmo da forma como aparece, assim como das informações que não foram apresentadas, trata-se de uma escolha tomadas no âmbito desta tese, para que obtivéssemos o produto final, que melhor atendesse ao consulente, sendo este um dos objetivos principais deste produto.

Como discutimos nos capítulos teóricos, muitos dicionários eletrônicos apenas reproduzem informações de dicionários impressos, que muitas vezes não são necessárias e nem atrativas para o usuário (L'HOMME E CORMIER, 2014). Neste sentido, conforme discutimos no capítulo 1 desta tese, procuramos selecionar as informações mais relevantes para apresentar ao usuário, sendo assim, ressaltamos que, para além dos aspectos que foram selecionados para compor o verbete, a base de dados – conforme apresentada no capítulo 2 – ainda armazena as seguintes informações associadas a cada UL:

i. formas flexionadas de todas as entradas, que permitem a identificação das ULs digitadas pelo usuário;
ii. anotações de sentenças para as camadas de elemento de frame, função gramatical e tipo sintagmático, que geram padrões de valência sintático-semântica utilizados na inferência de equivalentes de tradução;
iii. relações entre frames, que organizam semanticamente o repositório lexical da Diciopédia, ao mesmo tempo em que modelam os domínios do conhecimento cobertos pelo dicionário;
iv. *corpora*, de onde foram prospectados os termos e extraídas as sentenças para anotação;
v. tipos e gêneros textuais associados aos documentos que compõem os *corpora*, que contribuem para o controle de seu balanceamento e representatividade.

Mesmo considerando que todas essas informações são de extrema importância em nosso banco de dados e possuem dados relevantes para a constituição de ferramentas lexicográficas, nosso intuito é o de que esses dados estejam apenas no pano de fundo dando a estrutura necessária para que o consulente não-especialista obtenha dados satisfatórios a respeito de sua busca.

Desse modo, retomando a discussão que realizamos também no capítulo 1, desta tese a respeito dessa temática, temos observado cada vez mais que as ferramentas

lexicográficas têm se preocupado em proporcionar maior agilidade e objetividade para o usuário, e, neste processo, a teoria linguística tem deixado de apenas compor dados informacionais e tem estado cada vez mais presente na sustentação da interface das aplicações, ainda que não explícita, buscando atender especialistas e não-especialistas. Mesmo não se tratando de um produto para ser comercializado, já que se trata de um artefato acadêmico, tivemos essa preocupação no processo de síntese assim denominado por Atkins (2008). No próximo capítulo, iremos mostrar as análises de usabilidade desta função pelos usuários para verificarmos os aspectos teóricos em contraste com seu resultado prático no uso.

4.4 Análise dos aspectos lexicográficos concernentes ao aplicativo m.knob

Considerando as análises deste capítulo, a Tabela 6 mostra de maneira sintetizada as informações lexicográficas correspondestes ao aplicativo m.knob, assim como realizamos para as demais aplicações que foram descritas nesta tese.

PRÉ-LEXICOGRAFIA			
e-Dicionários	**Fontes de evidências**		
	Corpus objetivo	*Introspecção Subjetiva*	*Investigação Mista*
m.knob			
MACROESTRUTURA I – LISTAS DE PALAVRAS			
e-Dicionários	**Tipo**		
	Seletivo	*Abrangente*	
m.knob			
e-Dicionários	**Lista de Palavras-Núcleo**		
	Não-homográfica	*Parcialmente homográfica*	*Totalmente homográfica*
m.knob			
e-Dicionários	**Seleção de Conteúdo**		
	Tipo lexical	*Tipo morfológico*	*Tipo semântico*
m.knob			
MACROESTRUTURA II – ENTRADAS			
e-Dicionários	**Estrutura da Entrada**		
	Plana	*Em camadas*	
m.knob			
e-Dicionários	**Status da Entrada**		
	Critério da entrada principal	*Critério da sub-entrada*	
m.knob			
MICROESTRUTURA I – APRESENTAÇÃO			
e-Dicionários	**Diferenciação**		
	Plana	*Hierárquica*	
m.knob			
e-Dicionários	**Ordenação**		
	Histórica	*Por frequência*	*Básica*
m.knob			
MICROESTRUTURA II – TIPOS DE DADOS			

e-Dicionários	Interno					
	Forma				*Conteúdo*	
	Morf.	Orto.	Fon.	Lex.	Explicações	Tipos
m.knob						
e-Dicionários	**Externo I**					
	Sintagmático			*Paradigmático*		
	Comp.	Coloc.	Selec.	Gram.	Morf.	Sem.
m.knob						
e-Dicionários	**Externo II**					
	Uso				*Relação*	
	Reg.	Atual.	Estilo	Campo	Derivação	Ref. cruzada
m.knob						
e-Dicionários	**Outros**					
	Informativo				*Etimológico*	
m.knob						

Tabela 6:Resumo dos aspectos lexicográficos do m.knob

Fonte: Própria autora

Como se pode notar, o m.knob se assemelha, no que tange a esses critérios, aos e-dicionários EcoLexicon, Kicktionary, Field e Copa 2014. Na próxima seção, trataremos dos quantitativos de entradas lexicais por língua e dos aspectos que tangenciaram a escolha dos termos que compõem a Diciopédia.

4.5 Quantitativos de Entradas Lexicais por Língua

No total, a Diciopédia conta com 7.602 verbetes, assim distribuídos por língua: 2.370 para o Português Brasileiro; 3.180 para o Inglês e 2.052 para o Espanhol.

Apesar da discrepância em relação ao quantitativo de ULs em Espanhol ainda permanecer, tal diferença foi consideravelmente reduzida com o fruto do trabalho desta tese. A aplicação da metodologia descrita na seção 3.1 aumentou o número de termos específicos em Espanhol a partir de evidência em corpus. Os Gráficos 1 e 2 exemplificam como se deu esse processo, demonstrando-o para os primeiros 160 termos – em Português e em Espanhol, respectivamente – extraídos pelo TERMOSTAT.

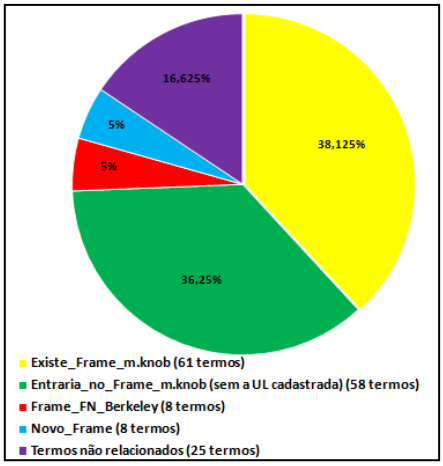


Gráfico 1: 160 primeiros termos gerados pelo TERMOSTAT para o Português.

Fonte: Própria autora.

A partir do Gráfico 1, verificamos que, para o Português, um percentual elevado de termos já estava cadastrado no aplicativo, conferindo legitimidade aos dados existentes. Foram acrescidos 58 novos termos ao m.knob e descartados 25 termos, que não foram condizentes com o domínio relacionado ao Turismo. Ademais, 8 termos foram acrescidos a novos frames criados e outros 8 termos foram adicionados a frames existentes da FN Berkeley.

Em relação à língua espanhola, por apresentar uma quantificação com maior defasagem no aplicativo, os resultados para a mesma quantidade em análise inicial (160 termos) demonstram que a maioria dos termos extraídos pelo TERMOSTAT não constava da base, conforme o Gráfico 2.

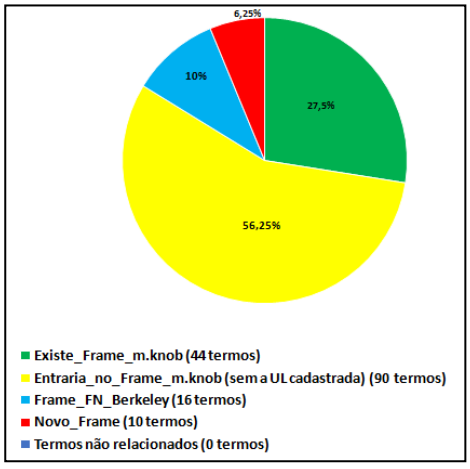


Gráfico 2: 160 primeiros termos gerados pelo TERMOSTAT para o Espanhol.

Fonte: Própria autora.

Dos 160 termos iniciais, 44 já existiam no aplicativo; 90 seriam acrescidos no m.knob; 16 termos foram adicionados aos frames existentes na FN Berkeley; 10 novos termos foram adicionados a novos frames e não obtivemos nesta análise inicial nenhum termo que não fosse relacionado ao Turismo.

Partindo da análise desses termos e da metodologia definida nesta tese, a tarefa de equiparar o quantitativo de unidades lexicais nas três línguas do m.knob foi repassada aos bolsistas de iniciação científica do projeto. Assim, buscou-se tornar a Diciopédia ainda mais eficiente para sua proposta, dando maior cobertura de pesquisa, assim como de dados sobre os domínios especializados, principalmente, no concernente ao domínio do Turismo, foco desta tese. Ademais, a prospecção de termos para um determinado domínio requer cuidado, devido aos empréstimos linguísticos e o limiar entre o que pertence ao campo mais genérico e ao mais específico.

Quanto ao número de verbetes do inglês, é importante ressaltarmos que, por termos os dados do inglês da FrameNet de Berkeley e esses poderem ser transferidos para a nossa base de acordo com os frames que evocam, um número maior de ULs nessa língua é esperado.

No que concerne às classes de palavras temos 131 adjetivos, 2 advérbios, 7.057 substantivos, e 412 verbos, considerando as três línguas. O número de substantivos é consideravelmente maior do que o das demais classes de palavras, uma vez que se trata de um dicionário terminológico. Entretanto, o recurso conta com um considerável número de verbos, o que se relaciona à abordagem teórica da Semântica de Frames.

No próximo capítulo, discutiremos os principais resultados obtidos relacionados ao uso da função Diciopédia.

5 ANÁLISE DE USABILIDADE DA FUNÇÃO DICIOPÉDIA DO MKNOB

Neste capítulo, descreveremos os resultados que foram apresentados tanto pela plataforma Appsee, como as respostas do questionário (ANEXO 1) que foram aplicados para cada um dos voluntários.

A aplicação desses testes tem como motivação investigar a relevância dos dados apresentados ao usuário segundo a proposta de verbete desenvolvida nesta tese. Para além de classificar o produto resultante conforme os critérios propostos por Atkins (2008), julgamos relevante averiguar de que modo o usuário interage com o dicionário eletrônico e como percebe sua interface. A partir dessas informações, poderemos alterar, acrescentar ou reduzir os elementos que compõem a Diciopédia, a partir das principais necessidades e exigências do consulente, construindo uma base empírica que sustente nossas decisões lexicográficas.

5.1 Análise do questionário de impressões de uso da função Diciopedia do Aplicativo m.knob

Trataremos, inicialmente, das análises de cada uma das perguntas realizadas no questionário. É importante ressaltarmos que a primeira parte do questionário correspondia aos dados pessoais do usuário, assim como o quesito referente ao termo de consentimento para participar como voluntário. Esses dados não serão discutidos aqui.

Para a primeira pergunta “Você costuma usar dicionários?”, obtivemos 92,1% de respostas sim (35 voluntários) e 7,9% de respostas não (3 voluntários), conforme o Gráfico 3. Esses dados mostram a importância do produto lexicográfico dicionário para a grande maioria dos voluntários.

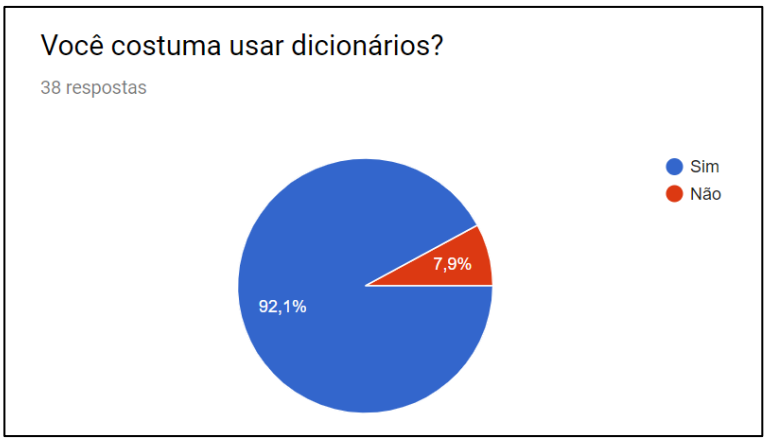


Gráfico 3: Resultado concernente ao percentual de pessoas que usam dicionários.

Fonte: Google Forms

A pergunta seguinte, de número 2, foi direcionada apenas para as pessoas que responderam sim na questão de número 1. Nesse sentido, apenas 35 (trinta e cinco) pessoas responderam essa pergunta. Conforme podemos ver no Gráfico 4, dos 35 voluntários, 33 (94,3%) afirmaram que acessam mais o dicionário eletrônico, enquanto 2 (5,7%) assinalaram acessar mais o de papel. Os resultados das duas primeiras perguntas conjugados atestam o êxito na seleção do corpo de voluntários, uma vez que a imensa maioria tinha tanto experiência com dicionários quanto com suas versões eletrônicas.

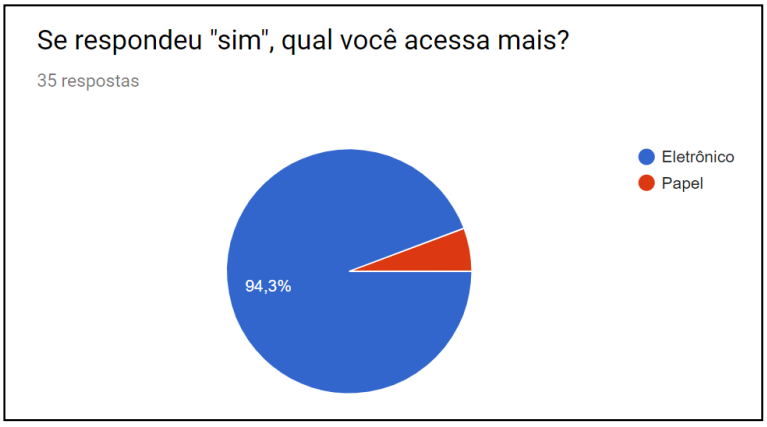


Gráfico 4: Resultado concernente ao percentual de tipo de dicionário mais acessado.

Fonte: Google Forms

A pergunta de número 3 está associada ao nível de proficiência na língua inglesa. Tal pergunta apresenta relevância para essa pesquisa, uma vez que apresentamos informações de tradução na Diciopédia, tanto para o inglês quanto para o espanhol.

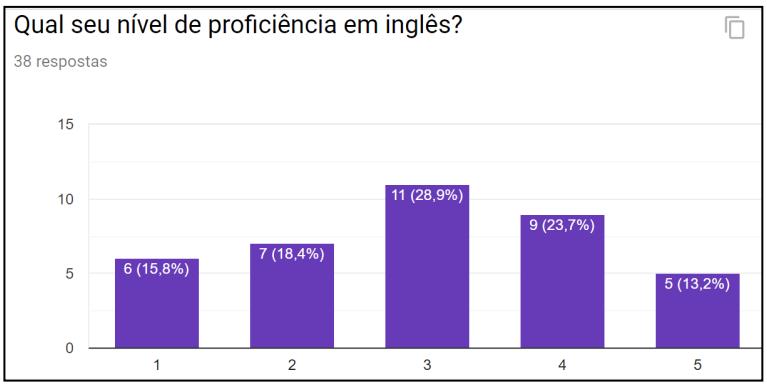


Gráfico 5: Resultado concernente ao percentual do nível de proficiência em inglês.

Fonte: Google Forms

Para essa pergunta, o usuário deveria selecionar seu nível de proficiência em uma escala de 1 a 5, sendo que, quanto maior o número, maior o nível de proficiência. De acordo com os dados apresentados no Gráfico 5, 6 voluntários (15,8%) marcaram o número 1, ou seja, o índice referente a não falar inglês. No nível 2, tivemos 7 voluntários (18,4%). Em relação ao nível médio, ou intermediário, tivemos 11 (28,9%), ou seja, a maior concentração de voluntários. No nível 4, tivemos 9 pessoas (23,7%), e, no nível mais alto de proficiência, detectamos 5 pessoas (13,2%), ou seja, a menor concentração.

A pergunta de número 4 se estruturou da mesma maneira, porém acerca da proficiência em espanhol. Assim sendo, obtivemos os seguintes resultados: 16 pessoas (42,1%) assinalaram o índice que revela o nível mais próximo a “não falar espanhol”, concentrando, portanto, grande parte dos voluntários. No nível 2, tivemos 12 voluntários (31,6%). No nível 3, intermediário, tivemos 6 (15,8%); no nível 4, identificamos 3 pessoas (7,9%); e, no nível mais avançado, apena 1 pessoa (2,6%), ou seja, a minoria.

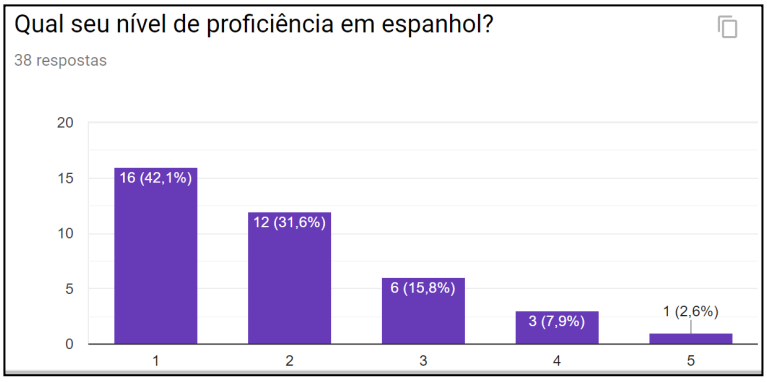


Gráfico 6: Resultado concernente ao percentual do nível de proficiência em espanhol.

Fonte: Google Forms

Depois de ter preenchido o formulário com esses dados iniciais, havia a seguinte orientação no próprio formulário:

> *Para responder as perguntas da próxima seção, será necessário que use a Diciopedia. Para tanto, você deve: 1. Clicar no ícone do m.knob na tela do celular que lhe será fornecido pela pesquisadora; 2. Acessar a função Diciopedia: dicionário temático dos campos do Turismo e dos Esportes; 3. Buscar as palavras de sua preferência e interagir com a aplicação da forma que achar mais conveniente; 4. Retornar à tela inicial do m.knob quando terminar, clicando nos balões localizados na parte inferior da tela da Diciopedia.*
> *ATENÇÃO: a função TRADUTOR não está sob avaliação e, portanto, não deve ser utilizada.*

Depois de ler essa orientação, os voluntários acessavam a Diciopédia, para, posteriormente, continuar avaliando a ferramenta por meio do questionário. Em nenhum momento, orientamos quanto ao tempo, ou número de palavras a serem pesquisadas. Deixamos o celular à disposição do voluntário, para que ele acessasse o quanto quisesse. Quando terminava de consultar, retornava ao questionário, dando sequência às perguntas seguintes. Foi durante esta seção do teste que os dados de usabilidade foram armazenados pelo Appsee.

Na pergunta subsequente, levando em consideração a seção de avaliação da Diciopédia, propusemos o seguinte enunciado "*Considerando os campos cobertos pelo m.knob (Turismo e Esportes Olímpicos), no geral, como você avalia os resultados das buscas pelas palavras desejadas?*". O Gráfico 7, aponta os resultados obtidos para essa análise, sendo que 36 usuários (94,7%) marcaram que encontraram resultado nas buscas realizadas e 2 (5,3%) assinalaram que o resultado encontrado não foi adequado à busca que tinham em mente. Nenhuma pessoa registrou a ocorrência de nenhuma palavra encontrada.

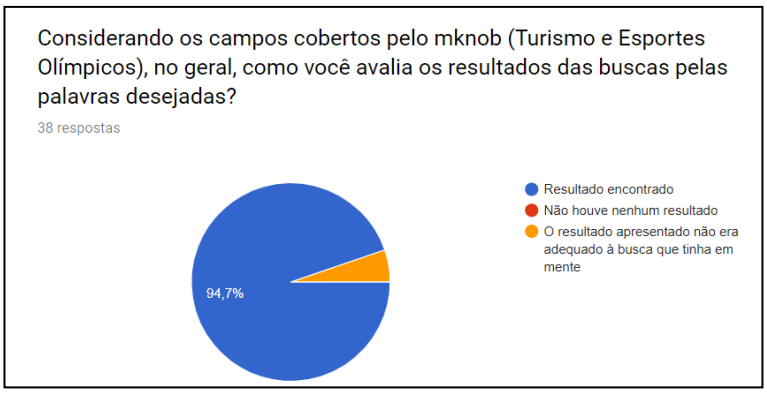


Gráfico 7: Resultado concernente ao percentual do nível de proficiência em espanhol.

Fonte: Google Forms

Na pergunta consecutiva, pedimos ao consulente que deixasse registrados quais foram os termos buscados que não obtiveram um resultado desejável. Conforme podemos ver no Gráfico 8, apenas 13 pessoas registraram algumas ocorrências. Excetuando-se o termo “medalha”, que 2 pessoas (15,4%) registraram não ter, todas as demais palavras sugeridas tiveram apenas 1 ocorrência de busca, totalizando um percentual de 7,7% correspondente a cada uma delas sendo: “balada”, “cavalo animal”, “estádio”, “jangada”, “fubá”, “ambrosia”, “taioba”, “angu”, “karatê”, “artes marciais”, “natação”, “vôlei”, “parque esportivo”, “vendedor ambulante”, “stand up”, “paralimpíadas”[61], “cutar” [62](custo),“dinheiro”, “pedalinho” e “torneio”.

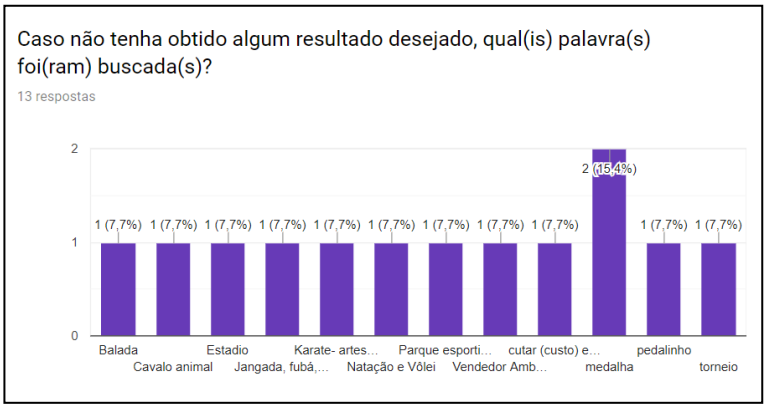


Gráfico 8: Resultado concernente ao percentual do nível de proficiência em espanhol.

Fonte: Google Forms

[61]Foi digitado assim.
[62]Foi digitado assim.

Após ter acesso a tais registros, e conhecendo o banco de dados, com o qual temos trabalhado ao longo desses últimos anos, estranhamos ter aparecido a ocorrência de determinadas palavras, como “estádio”, “natação”, “vôlei”, “parque esportivo” e “custar”, já que são informações que estavam cadastradas em nosso banco de dados. Dessa forma, não sabemos se a informação apresentada não correspondeu ao que a pessoa buscava, ou se, porventura, por problemas de digitação, houve alguma falha neste sentido.

As demais que foram assinaladas, como “balada”, “cavalo animal”, “jangada”, “fubá”, “ambrosia”, “taioba”, “angu”, “karatê”, “artes marciais”, “vendedor ambulante”, “stand up”, “paralimpíadas”, “dinheiro”, “pedalinho” e “torneio”, não obtiveram resultado encontrado. Ressaltamos, aqui, que o termo “cavalo” foi especificado como animal, uma vez que há vários resultados com o termo “cavalo”, conforme mostra a Figura 70, que se relacionam, principalmente, com o domínio dos Esportes.

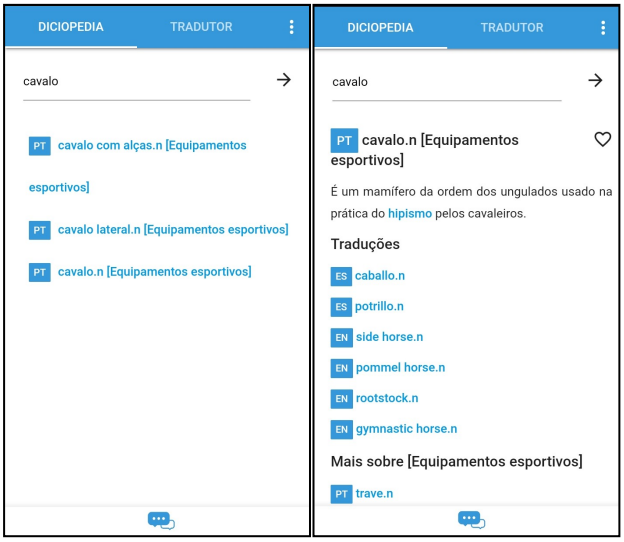


Figura 70: Resultado de busca do termo “cavalo” na Diciopédia

Na sequência do questionário, temos a seguinte pergunta “Você chegou a sugerir este (s) termo (s) no aplicativo?”, e como resultado – Gráfico 9 –, das 23 pessoas que responderam a essa pergunta, 17 delas (73,9%) alegaram não ter sugerido. As outras 6 (26,1%) disseram ter sugerido, ou seja, a minoria. Tal resultado demonstra que, apesar do grupo de voluntários estar, em sua maioria, acostumado a utilizar dicionários eletrônicos, isso não garante que eles enxerguem as possibilidades dinâmicas que tal plataforma traz à confecção de produtos lexicográficos. Uma razão para isso pode estar

exatamente no fato de que, em sua maioria, conforme discutido no capítulo 1, os e-dicionários são versões eletrônicas de dicionários de papel.

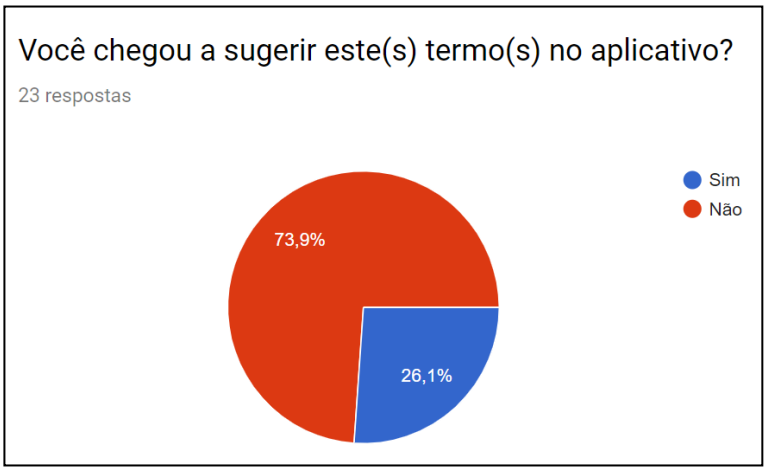


Gráfico 9: Resultado concernente ao percentual de termos que foram ou não sugeridos.

Fonte: Google Forms

No que tange ao nível de facilidade em utilizar a Diciopédia, também realizamos uma pergunta com base nos mesmos parâmetros utilizados para as perguntas sobre proficiência em línguas, ou seja, a ser respondida em uma escala de 1 a 5: quanto menor o número, mais difícil seria de usá-la, e quanto mais próximo ao número 5, mais fácil.

De acordo com o Gráfico 10, 29 pessoas (76,3%) alegaram ser muito fácil, outros 6 (15,8%) consideraram o nível 4 e 3 (7,9%) consideraram ser de nível intermediário. Assim, nota-se que, de um modo geral, a interface proposta não apresentou dificuldades ao usuário, o que é um indicador muito positivo.

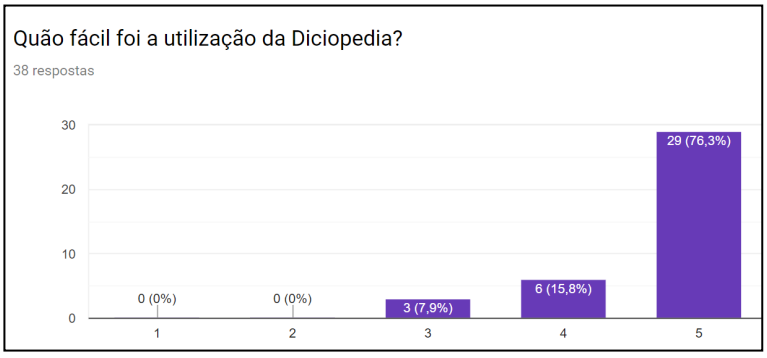


Gráfico 10: Resultado concernente ao percentual do nível de facilidade em acessar a Diciopédia.

Fonte: Google Forms

A próxima pergunta do questionário remete à avaliação da funcionalidade do aplicativo, tendo em vista o público-alvo a que se destina, que, neste caso, são os turistas. Também consideramos uma escala de 1 a 5, que varia de totalmente inútil (1) a muito útil (5). Com base no Gráfico 11, 26 voluntários (68,4%) expressaram ser muito útil; no nível 4, tivemos 7 pessoas (18,4%); e no nível 3 (intermediário) foram 5 usuários (13,2%). Novamente, consideramos o resultado muito positivo, uma vez que ele aponta para uma avaliação pós-uso livre do aplicativo.

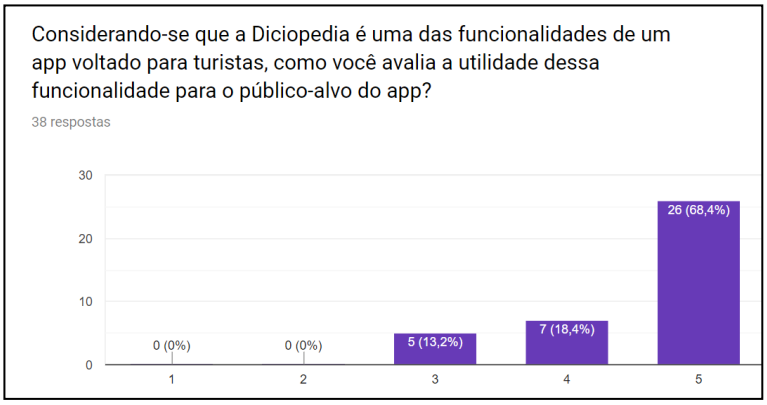


Gráfico 11: Resultado concernente ao percentual do nível de utilidade da Diciopédia para o público-alvo.

Fonte: Google Forms

A etapa que se segue faz alusão às informações concernentes às traduções apresentadas pela Diciopédia. Pergunta-se se o voluntário considera as traduções uma informação importante na Diciopédia e os resultados obtidos foram os seguintes, segundo consta no Gráfico 12: 36 consulentes (94,7%) afirmaram ser importante; 1 pessoa (2,6%) alegou que não é importante e 1 (2,6%) não reparou na existência das traduções.

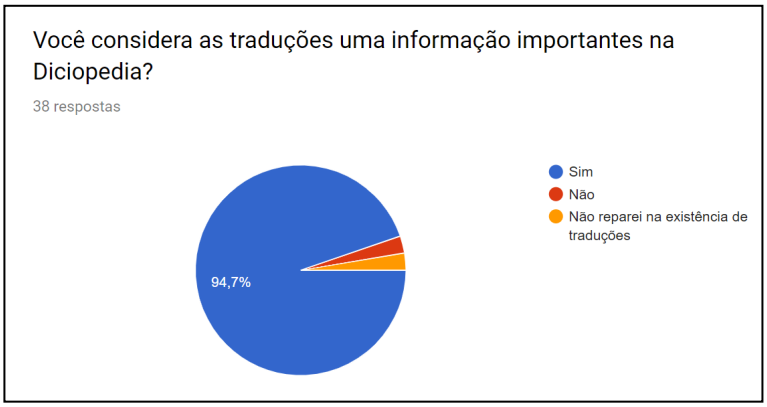


Gráfico 12: Resultado concernente ao percentual de importância das traduções na Diciopédia.

Fonte: Google Forms

Esse dado ganha relevância especial se considerado o fato de que a maioria dos participantes alegou não ter elevados níveis de proficiência nas línguas estrangeiras alvo do aplicativo, o que, a princípio, poderia reduzir seu interesse por traduções.

Na sequência, a pergunta foi concernente a como o usuário avalia as traduções apresentadas pela Diciopédia. Neste quesito, apenas 37 pessoas responderam, uma vez que uma das pessoas marcou que não havia reparado na existência de traduções. O critério de análise também foi posto em uma escala de 1 a 5, sendo que o número 1 representa traduções muito ruins e o número 5 representa traduções muito boas. Conforme mostra o Gráfico 13, 28 pessoas (75,7%) avaliaram as traduções como muito boas; 7 delas (18,9%) alegaram ser de nível 4 e 2 (5,4%) assinalaram ser de um nível médio. Apesar de o nível de proficiência da maioria dos participantes não ser alto, a percepção deles sobre a adequação das traduções foi, o que pode ser devido ao fato de eles terem buscado palavras cujas traduções conheciam, ou mesmo por terem acreditado que as traduções estavam corretas, mesmo não as conhecendo, o que não seria diferente da experiência de um usuário não-proficiente com qualquer outro dicionário.

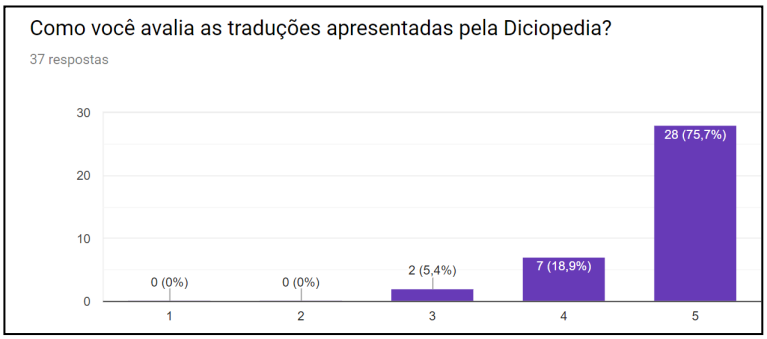


Gráfico 13: Resultado concernente ao percentual das avaliações das traduções na Diciopédia.
Fonte: Google Forms

A próxima seção de perguntas é referente aos verbetes da Diciopédia. É importante ressaltarmos que em cada uma das perguntas dessa seção foram mostradas telas de exemplos de busca e de resultados obtidos concernentes ao termo “feijoada.n”, que utilizamos, inclusive, como ilustração no capítulo 4 desta tese. Colocamos tais telas de exemplos, de modo que a pessoa seja lembrada ou que entenda sobre o que se trata a pergunta exatamente.

Prosseguimos com a seguinte indagação: “*qual a relevância da informação apresentada entre colchetes ao lado da lista de resultados de uma pesquisa?*”. Tal

pergunta buscava avaliar a relevância de apresentar o nome do frame evocado pela UL ao consulente. Observamos, no Gráfico 14, que 32 participantes (84,2%) consideraram o nível de importância como sendo 5, ou seja, muito; 4 delas (13,2%) assinalaram o nível 4 e apenas 1 pessoa (2,6%) marcou o nível 3. Nenhuma pessoa marcou os níveis 1 ou 2, o que indicaria ser esta uma informação irrelevante. Acreditamos que o fato de os nomes de frames e EFs na FrameNet serem mnemônicos (RUPPENHOFFER ET AL. , 2016) contribui para a percepção de sua relevância pelo usuário não-especialista.

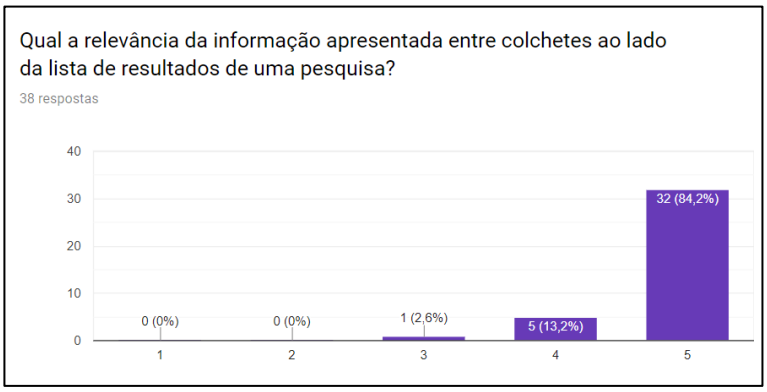


Gráfico 14: Resultado concernente ao nível de relevância dos nomes dos frames nos resultados encontrados na Diciopédia.

Fonte: Google Forms

Ainda sobre os verbetes, a pergunta subsequente diz respeito ao fato de os participantes terem clicado ou não em algum dos hiperlinks presentes nas definições dos termos. Considerando o Gráfico 15, 31 participantes (81,6%) afirmaram tê-lo feito, e 7 (18,4%) marcaram que não.

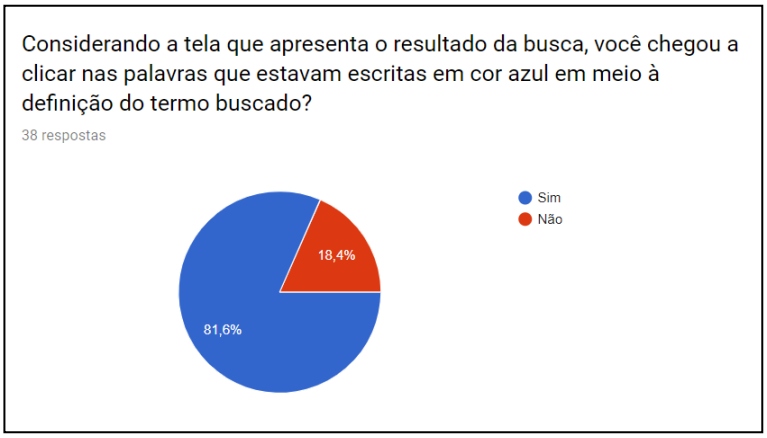


Gráfico 15: Resultado concernente ao nível de relevância dos hiperlinks presentes na Diciopédia.

Fonte: Google Forms

O item seguinte pergunta sobre a relevância das informações que foram acessadas ao se clicar nos hiperlinks, o que denominamos de palavras em cor azul. Das 38 pessoas consultadas, 25 (65,8%) afirmaram que são muito relevantes (nível 5), 12 participantes (31,6%) marcaram o nível 4 e apenas 1 (2,6%) sinalizou como médio. Os pontos 1 e 2 da escala indicam irrelevância, situação em que não houve ocorrência de respostas, conforme mostra o Gráfico 16.

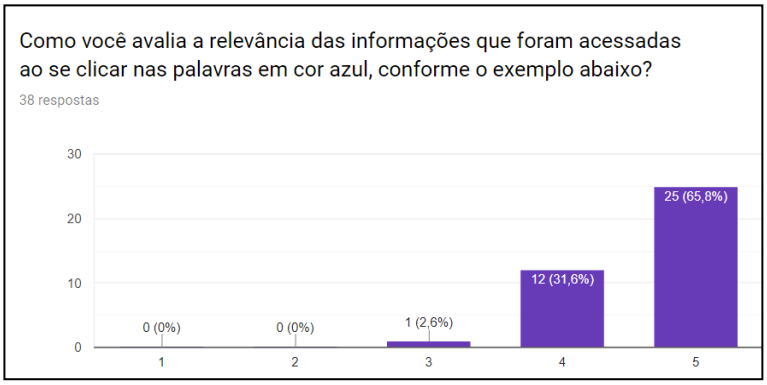


Gráfico 16: Resultado concernente ao nível de relevância dos hiperlinks presentes na Diciopédia.

Fonte: Google Forms

Conjugados, os resultados mostrados nos Gráficos 15 e 16 apontam para o fato de que os hiperlinks foram ferramenta eletrônica bem explorada pelos consulentes, o que sugere ser esta uma funcionalidade típica dos e-dicionários mais internalizada pelo usuário, diferentemente da ferramenta colaborativa de sugestão de palavras. Tal fato pode estar ancorado na ideia de que, ao contrário do que ocorre com a navegação por hiperlinks, em que o consulente se mantém nesse papel, o de visualizador de informações, nas ferramentas colaborativas ele precisa assumir um papel de autor do produto lexicográfico, o que parece representar ainda uma barreira na experiência do usuário.

Posteriormente, procuramos avaliar as estruturas qualia que foram inseridas abaixo da definição dos termos na Diciopédia e, neste sentido, também procuramos saber qual o nível de relevância delas para o consulente. Como uma forma de parametrizar as perguntas, mantivemos a escala de 1 a 5, novamente com os mesmos critérios anteriores, ou seja, 1 “totalmente irrelevante” e 5 “muito relevante”. O Gráfico 17 mostra que 25 consulentes (71,1%) afirmaram ser muito relevantes; 10 (26,3%) consideraram o nível 4 e 1 (2,6%) assinalou o nível médio, ou seja, 3. Dessa forma, as relações qualia ternárias propostas por DINIZ DA COSTA (em preparação) e PAIVA (em preparação), a

princípio com foco na melhoria do desempenho do sistema de recomendação (vide DINIZ DA COSTA ET AL., 2018), mostraram-se também relevantes na função Diciopédia do m.knob.

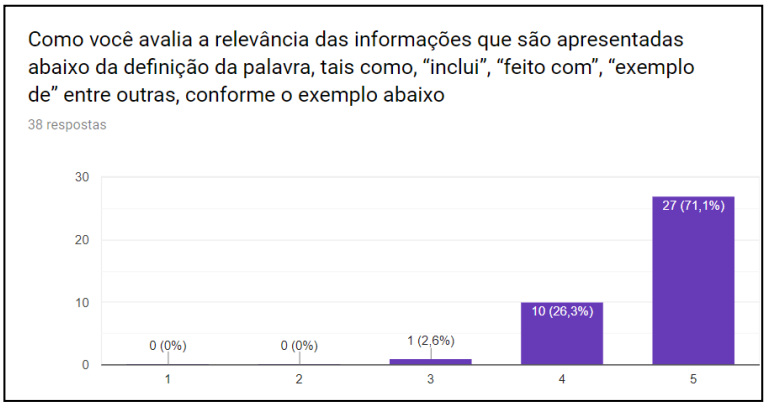


Gráfico 17: Resultado concernente ao nível de relevância das estruturas qualia presentes na Diciopédia.

Fonte: Google Forms

O próximo item se refere à lista de palavras que são listadas na opção “Mais sobre”, que, conforme apontamos anteriormente, lista as palavras que evocam o mesmo frame evocado pela palavra-chave do verbete. A primeira indagação a respeito dessas informações foi se o participante chegou a acessar algum desses termos sugeridos. Como resultado, obtivemos que 29 pessoas (76,3%) disseram que sim e 9 (23,7%) marcaram a opção “não”, conforme mostra o Gráfico 18.

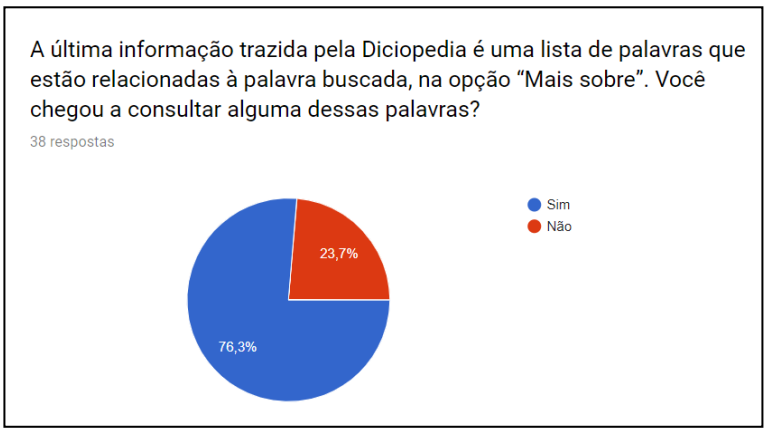


Gráfico 18: Resultado concernente ao acesso das ULs que compõem o mesmo frame na Diciopédia.

Fonte: Google Forms

Ainda sobre essa lista de palavras apresentadas na Diciopédia, a segunda pergunta foi “*Essas palavras listadas ampliaram o seu leque de busca?*”. Para esse item tivemos 34 usuários (89,5%) que afirmaram que sim e 4 deles (10,5%) consideraram que “não”, segundo consta no Gráfico 19.

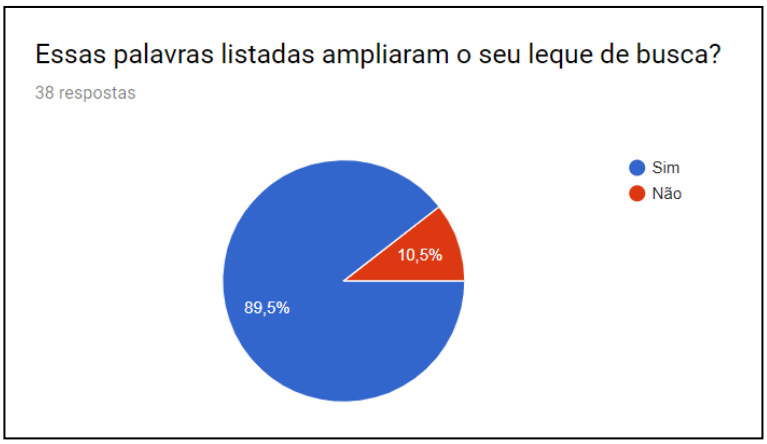


Gráfico 19: Resultado concernente à lista de ULs do mesmo frame quanto ampliar o leque de busca na Diciopédia.

Fonte: Google Forms

O último item sobre a lista de palavras procurou averiguar se ela poderia auxiliar na definição da palavra relacionando-a em uma rede de significados e, conforme mostra o Gráfico 20, todos os participantes alegaram que “sim”, totalizando 100%.

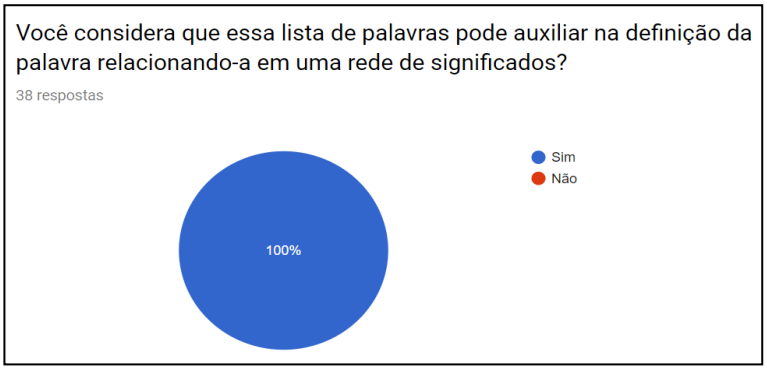


Gráfico 20: Resultado concernente ao aspecto de relacionar a definição a uma rede de significados.

Fonte: Google Forms

Tomados em conjunto, os resultados compilados pelos Gráficos 18, 19 e 20 apontam que a organização do léxico em termo de frames pode ser positiva para o consulente não-especialista, na medida em que ela motiva novas consultas, amplia o

vocabulário e auxilia na definição do sentido do item pesquisado (para uma discussão do papel dos frames na organização sob outro viés, vide DA SILVA E CHISHMAN, 2018).

A penúltima pergunta listada no questionário em análise foi sobre a avaliação dos participantes quanto às informações dadas pela Diciopédia. E para tal, realizamos uma nova escala de 1 a 5, sendo 1 incompletas até chegar ao nível 5 muito completas. Neste sentido, 25 pessoas (65,8%) assinalaram ser muito completas, 11 (28,9%) escolheram o nível 4 e apenas 2 (5,3%) indicaram o nível 3, ou médio, segundo consta no Gráfico 21.

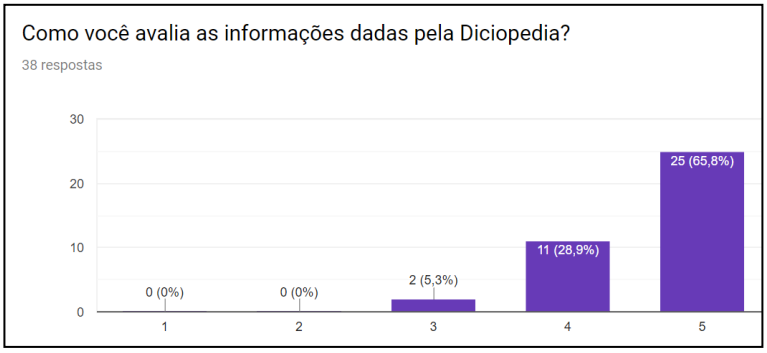


Gráfico 21: Resultado concernente à avaliação das informações trazidas pela Diciopédia.
Fonte: Google Forms

Por último, e não menos importante, pedimos que nos informassem o seguinte: *"se você considera incompletas as informações dadas pela Diciopédia, que tipo de informação ou dado você incluiria?"*. Neste quesito, apenas 10 voluntários se expressaram por escrito, conforme Figura 71.

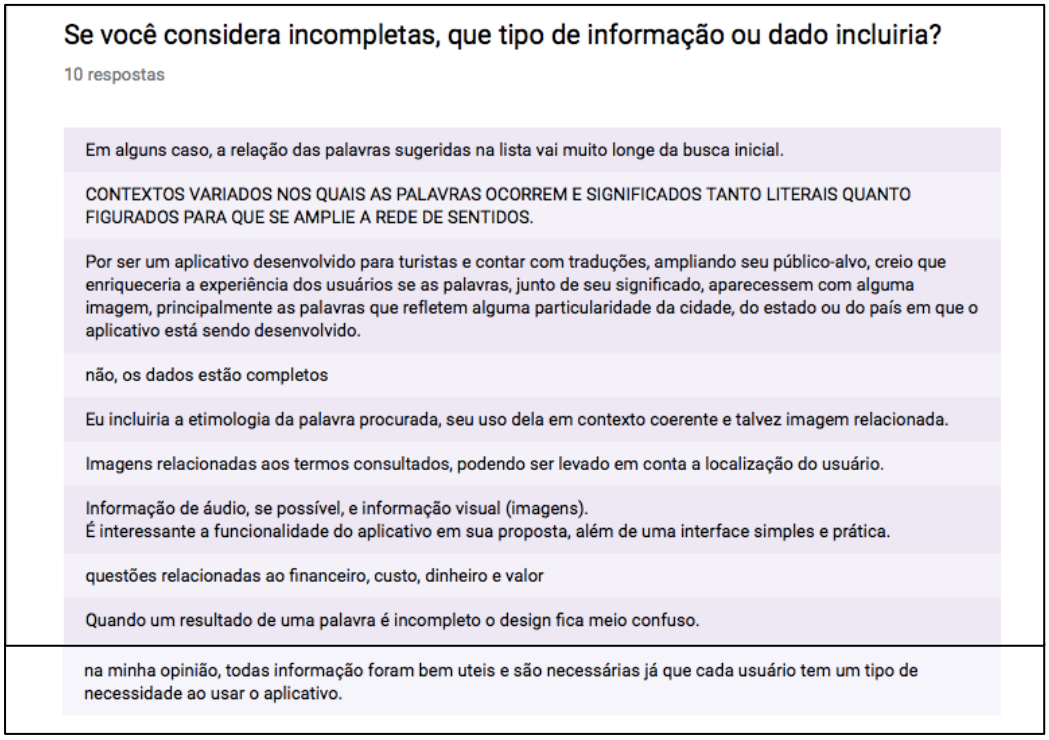


Figura 71: Resultado a respeito de dados o informações sugeridas pelos usuários para a Diciopédia.
Fonte: Google Forms

Com base nesses registros, e nos dados fornecidos pela plataforma Appsee, nosso intuito é propor alterações para a Diciopédia, no último capítulo desta tese, de modo a atender as reais necessidades do consulente. Mesmo que não seja implementado de imediato, trabalhos futuros e posteriores poderão dar continuidade a esta investigação.

Na próxima seção, trataremos dos dados capturados pela plataforma Appsee.

5.2 Análise dos dados fornecidos pela plataforma Appsee

Conforme descrevemos no capítulo 3, de metodologia, os dados de interação das sessões realizadas pelos 38 voluntários foram armazenados pelo Appsee. Ressaltamos mais uma vez que nenhum dado pessoal do usuário foi armazenado pela plataforma, sendo que todas as interações com o aplicativo foram realizadas em um único telefone celular, fornecido pela pesquisadora.

Com base no número de participantes, a plataforma fez uma amostragem randomizada de 30 sessões e faremos nossas análises com base nesses dados fornecidos.

Iniciaremos analisando, a título de exemplo de como se deu o teste, um dos vídeos de acesso ao aplicativo, disponível em https://drive.google.com/open?id=190U3SjNN86K-Zjvn2yp86uCqUhayKueW.

O vídeo, com duração de 8:42 minutos, registrado como *session* 18, revela a validação da aplicação do questionário de uso da Diciopédia, a partir do qual os participantes acessavam inicialmente as orientações desse teste, e, depois de preencherem o formulário com os dados iniciais, é que começavam a usar o aplicativo. Tal fato pode ser constatado no vídeo, uma vez que o aplicativo foi aberto e, apenas depois de passado 1:11 minuto, é que obtivemos a sequência de navegação no aplicativo.

O primeiro termo a ser buscado foi vôlei, conforme mostram as imagens do vídeo. As opções que apareceram para essa busca foram as seguintes: jogador de vôlei.n (`Atletas_por_esporte`); vôlei de praia.n (`Esportes`) e vôlei.n (`Esportes`) no registro de quando havia se passado 1:26 minuto. A opção escolhida pelo usuário foi a última descrita, vôlei.n, em 1:28 minuto. Possivelmente, o participante leu as informações que estavam sendo fornecidas pela Diciopédia de 1:28 até 1:46, já que durante esse tempo a tela ficou inerte.

Uma nova busca foi realizada pelo termo “ibero”, e nenhum resultado foi encontrado para essa pesquisa. Em uma nova tentativa de busca, o usuário digitou o termo “abertura”, quando já estamos em 2:07 minutos de gravação. Os resultados exibidos foram: abertura.n (`Cerimônia`); abertura.n (`Jogadas_pontuadas`) e abertura.n (`Jogadas_individuais`). No tempo de 2:13 minutos, o termo acessado foi “abertura”.n (`Cerimônia`).

Passados alguns segundos, em 2:20 minutos, foi realizada uma nova pesquisa com o termo “medalha”, e nenhum resultado foi obtido, e, posteriormente, buscou-se o termo “ganhador”, que também constou como nenhum resultado encontrado. Nesses casos em que não há um resultado obtido, havia a possibilidade de sugerir esse termo e nenhum deles foi sugerido. Essa informação deve ser alinhada com a pergunta que realizamos no questionário de avaliação do uso da Diciopédia, no qual o Gráfico 9 mostra que 73,9% dos participantes não sugeriram os termos não encontrados no m.knob, ou seja, a maior parte dos consulentes, sendo que apenas 26,1% disseram tê-lo feito.

No tempo marcado em 2:48 minutos, o participante digitou o termo “ouro”, obtendo como resultado ouro.n (`Desejabilidade`), acessando para ver a definição. Em 3:12 minutos realizou uma nova busca pelo termo “rio de janeiro”, e a Diciopédia listou os

seguintes itens: “rio.n” (Locais_naturais) tanto para o português, quanto para o espanhol. Como não foi o resultado esperado, o participante apagou tal termo e, no tempo de 3:32 minutos, efetuou a busca por “Grécia”, não obtendo nenhum resultado para esta busca, uma vez que, assim como no caso anterior, tratava-se de nome próprio, não coberto, portanto, pelo dicionário. Também não sugeriu o termo.

Posteriormente, quando o tempo de teste já marcava 3:40 minutos houve uma nova pesquisa pelo termo “taekwendo” e, como retorno, a Diciopédia trouxe os resultados: “taekwendo.n” (Esportes) tanto para o português quanto para o inglês. O consulente clicou no resultado do português no tempo de 3:54 minutos e, possivelmente, depois de efetuar a leitura buscou um novo termo aos 4:02 minutos, sendo esse “judô”, e, como resultado, tivemos “judô.n” (Esportes). Houve o clique na opção disponibilizada em 4:09 minutos e, na sequência, o participante rolou a tela e foi lendo as opções mostradas em “Mais sobre”, clicando sobre o termo “badminton.n” aos 4:25 minutos. Em seguida, rolou mais a tela do celular, buscando as opções oferecidas em “Mais sobre”, clicando em “maratonas aquáticas.n” (Esportes). Novamente, seguiu rolando a tela e clicou em “pentatlo moderno.n” (Esportes). Esses dados exemplificam as respostas apresentadas nos Gráficos 19 e 20, que, respectivamente, perguntavam se essa lista ampliava o leque de busca e se tal listagem de palavras poderia auxiliar na definição da palavra, considerando uma rede de significado.

Quando o tempo do teste estava em 5:04 minutos, o voluntário clicou para voltar para a tela inicial, instrução dada no questionário de uso, para encerrar a sessão. Retornou novamente para o ícone Diciopédia em 5:07 minutos e a tela estava no último resultado exibido, (pentatlo moderno), voltou a fazer a rolagem de barra na tela, buscando outros termos. Clicou na UL “tênis de mesa” aos 5:28 minutos de acesso; logo em seguida, clicou em “lutas.n” (Esportes) e retornou à tela inicial do aplicativo aos 5:40 minutos. O aplicativo ficou aberto nesta tela de entrada até 6:01 minutos, sendo que imediatamente retornou para a tela de busca, na qual constava o último resultado de “lutas.n”. O aplicativo ficou aberto nesta tela sem nenhum movimento até os segundos finais de acesso desse participante, quando o tempo já estava em 8:39 minutos, quando retornou para a tela inicial para encerrar a sessão em 8:42 minutos.

O vídeo gravado exemplifica a jornada do usuário no aplicativo, mas, para fins do teste de usabilidade, nos interessam as tendências reveladas pela sobreposição de todas as jornadas em uma amostra representativa. Cada toque, deslize e ação de cada

tela foi gravado, o que nos permitiu obter uma avaliação da experiência do usuário com a Diciopédia. É sobre esses dados que nos debruçamos nas seções seguintes.

5.2.1 Estatísticas de Visualização de Telas pelos Usuários

O primeiro conjunto de dados apresenta uma visão geral da usabilidade do aplicativo, organizada pelas estatísticas de visualização de cada tipo de tela da Diciopédia. Elas apontam para a validade do teste realizado, dado o fato de demonstrarem que os voluntários seguiram adequadamente as instruções fornecidas.

Conforme temos na Figura 72, a aplicação Appsee nos forneceu o seguinte ranqueamento de interação dos usuários em relação ao tempo de maior uso em cada uma das seguintes telas: (i) termo Diciopédia (61,3%); (ii) Diciopédia (acesso) (16,5%), tela do resultado de busca (10,3%) e na tela de entrada do aplicativo (8,9%).

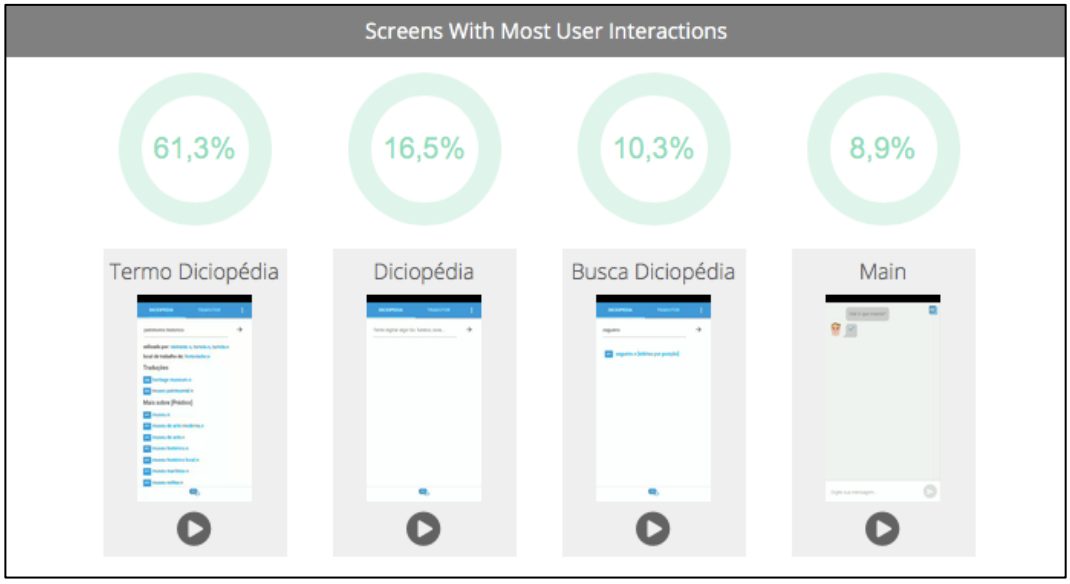


Figura 72: Resultado do percentual de interação do usuário com as telas do experimento.
Fonte: https://www.appsee.com/

Esses dados gerados revelam que o consulente seguiu adequadamente as instruções no questionário de uso e que alcançamos o objetivo de interação do usuário com o fator primordial de análise, ou seja, os dados dos termos da Diciopédia, o que valida o experimento.

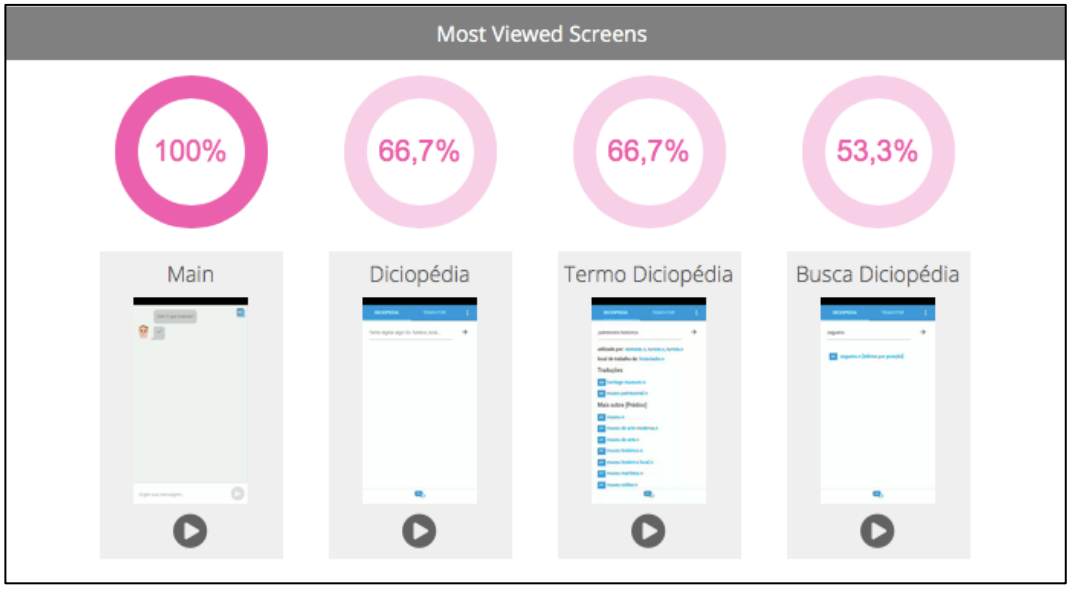


Figura 73: Resultado do percentual das telas mais visualizadas da Diciopédia.

Fonte: https://www.appsee.com/

Na Figura 73, onde temos o registro das telas mais visualizadas, podemos observar que a tela denominada “main’, ou seja, a tela principal, de entrada do aplicativo, teve o maior percentual (100%), e isso se deve a três fatores principais: primeiro, a tela de acesso era a primeira a ser aberta durante o teste; segundo, orientamos os usuários a retornar para a página principal no momento que finalizassem sua busca, de modo a encerrar sua sessão de uso, e, terceiro, foi a partir desta tela que o Appsee fez o recorte inicial da amostragem.

As telas seguintes onde digitamos a busca e o termo Diciopédia alcançaram o mesmo percentual de visualização (66,7%), tal aspecto se justifica pelo fato de o Appsee fazer um recorte amostral de 20 usuários para a compilação dos dados acerca das telas com grande fluxo de visitação. Isso é feito para otimizar o uso de espaço de armazenamento e não compromete os resultados do teste, uma vez que 20 é o número de usuários requerido para os mais detalhados testes de usabilidade, conforme Alroobaea e Mayhew (2014).

Já a última tela da Figura 73, a tela de busca foi carregada menos vezes que as demais devido ao fato de os usuários terem repetidamente prosseguido sua interação com o aplicativo a partir dos hiperlinks disponibilizados nos verbetes, o que pode ser confirmado pelos resultados apresentados no 15.

A última das estatísticas de visualização de telas diz respeito ao engajamento dos usuários. O engajamento é medido pelo tempo médio que cada usuário passou em cada tela visitada. Esse dado é mostrado na Figura 74.

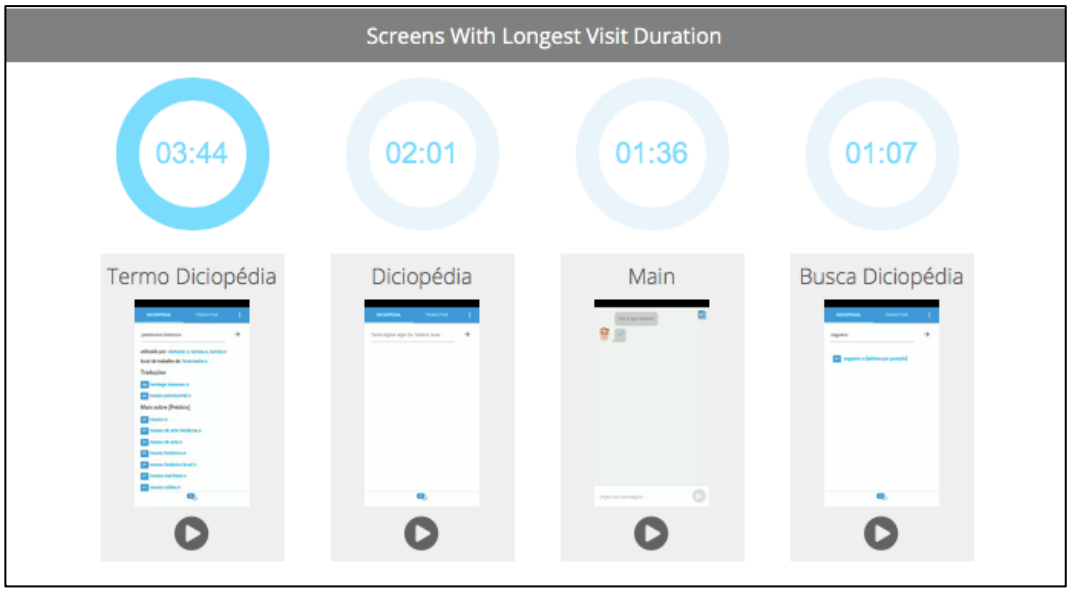


Figura 74: Resultado das telas com maior tempo de duração de visita médio da Diciopédia.
Fonte: https://www.appsee.com/

A tela que apresentou maior duração média na visita foi a do “termo da Diciopédia”, com o registro de 03:44 minutos em média por sessão. Esse dado corrobora as análises anteriores, segundo as quais, por ser a tela com os dados principais da Diciopédia, por conter o maior número de informações e por ser uma tela que exige um engajamento cognitivo maior, os participantes se detiveram mais diante dela. Em seguida, tivemos a tela da Diciopédia, local onde se efetuaram as buscas com duração média de 02:01 minutos, o que pode ser justificado pelo fato de o consulente precisar digitar o termo a ser pesquisado e até mesmo pensar qual termo irá buscar, indicando também uma interação maior. É importante considerarmos que, por ser um dicionário de domínio específico, e os usuários foram informados sobre isso, tiveram que pensar sob essa perspectiva. Tratando-se de um usuário que já saiba da utilização e dos domínios fornecidos, pode ser que este tempo não se repita nesse contexto. Por conseguinte, temos a tela principal com o registro de 01:36 minuto, e, por último, a tela com o resultado obtido da busca na Diciopédia, com 01:07 minuto.

Na próxima seção, demonstraremos por meio dos gráficos elaborados pela aplicação Appsee, o principal fluxo de usuários para cada um dos percursos e telas da Diciopédia.

5.2.2 Fluxo de Usuários

Nesta seção, apresentaremos os dados relativos ao fluxo dos usuários durante o uso do aplicativo m.knob. Tais fluxos são sempre definidos a partir de uma tela de referência, para qual são quantificados os principais caminhos trilhados durante a utilização do aplicativo.

De acordo com a Figura 75, que representa o fluxo de usuários na tela principal do m.knob – "main" – foram registradas 30 sessões – aleatoriamente escolhidas pelo Appsee para a amostragem – e o mesmo número de impressões, ou seja, de aberturas da tela com interações distintas. Para este percurso foi constatado 1 fluxo para a tela de entrada (*inbound screen*), indicando que todos as sessões da amostra chegaram à tela principal após a abertura do aplicativo, conforme as instruções constantes no questionário. Já para as telas de saída (*outbound screens*), foram 4 possibilidades: 56,7% para a Diciopédia, 30% para o fechamento do aplicativo (seguindo a instrução ao final da seção 2 do questionário), 10% para a tela de resultados da Diciopédia, caso o usuário tivesse clicado nos balões de diálogo para voltar à tela principal e, na sequência, clicado sobre o ícone da Diciopédia novamente e apenas 3,3% que, desobedecendo as instruções ou não as compreendendo corretamente, clicaram na função do tradutor. O fluxo de navegação apresentado novamente reforça a correta compreensão, pelos participantes, da proposta experimental.

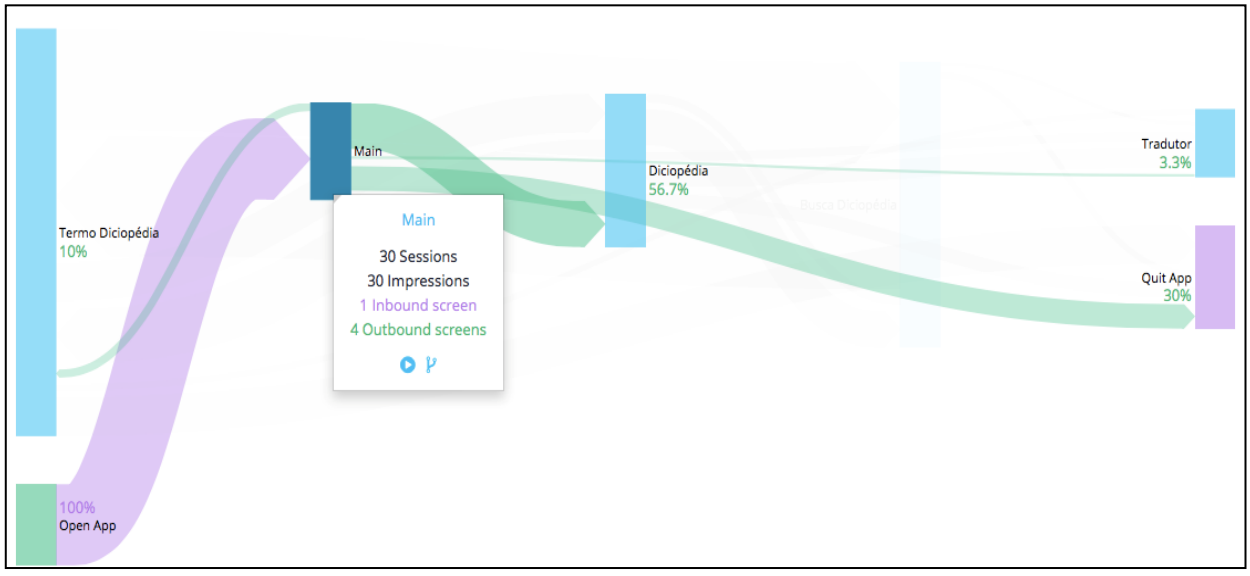


Figura 75: Resultado do fluxo do usuário para a tela "Main" (principal) da Diciopédia.
Fonte: https://www.appsee.com/

Para a tela de chegada da Diciopédia, a Figura 76 revela que foram registradas 20 sessões, sendo 41 impressões, ou seja, aberturas da tela com interações distintas. Obtivemos 4 fluxos relacionados às telas de entrada (*inbound screen),* sendo que 34,1% clicaram da tela do "Termo Diciopédia" para a "Diciopédia"; 41,5% se deslocaram da tela principal "Main" para a tela da Diciopédia; 17% se moveram da tela Diciopédia para o tradutor, e um pequeno rastro sem percentual mostra o deslocamento da tela "Busca Diciopédia" para a tela da "Diciopédia". Essa última situação deve-se, principalmente, ao fato de que, para realizar uma nova busca, basta apagar a informação contida no campo de busca para acessar uma nova informação.

Já para as telas de saída (*outbound screens*), tivemos também 4 possibilidades: 48,8% se deslocaram da tela "Diciopédia" para a tela "Termo Diciopédia", o que demonstra que sucessivas buscas foram realizadas, uma vez que o percentual maior aparece nesse sentido. Também o percentual de 31,7% entre a tela "Diciopédia" e a tela "Busca Diciopédia" corrobora esses índices de sucessivas buscas. Tivemos também o registro de 7,3% migrando para a tela do "Tradutor", fator esse que poderia ser minimizado com a retirada dessa função que está ao lado de Diciopédia, conforme podemos ver na tela "Diciopédia". E, por último, temos a obtenção de 13,2% que saíram da tela "Diciopédia" para fechar o aplicativo.

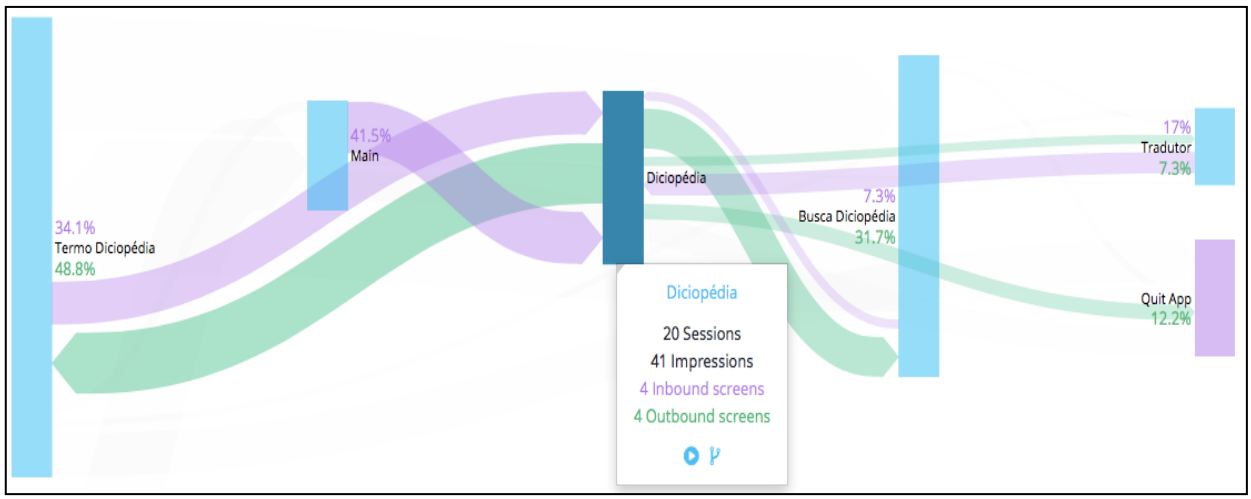


Figura 76: Resultado do fluxo do usuário para a tela da Diciopédia.

Fonte: https://www.appsee.com/

A Figura 77 comprova que os voluntários efetivaram inúmeras buscas e visualizaram termos em idas e vindas. Tal aspecto reforça a ideia de que os usuários se engajaram no uso da função, o que pode ser indicativo do interesse gerado pela interface

e sua facilidade de uso, uma vez que as instruções no questionário não instruíam quanto à realização de sucessivas buscas.

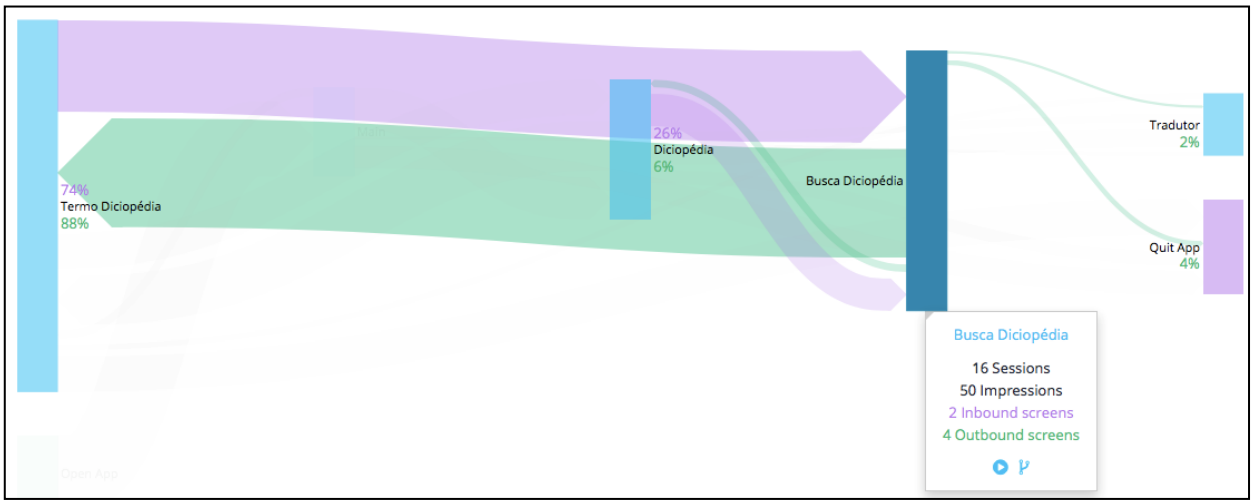


Figura 77: Resultado do fluxo do usuário para a tela de busca da Diciopédia.
Fonte: https://www.appsee.com/

A Figura 78 também corrobora o fluxo do usuário entre as telas do “termo diciopédia” e “busca diciopédia”, demonstrando que foram muito recorrentes. Em 20 sessões, que resultaram em 67 impressões, obtivemos os maiores fluxos entre a tela “Busca diciopédia” e “termos diciopédia” com 65,7% (*inbound screen),* enquanto o processo inverso resultou em 55,2% (*outbound screens*). Considerando a tela “Termo Diciopédia”, 20,9% fizeram esse percurso para sair do aplicativo, e apenas 3% saíram dessa tela para a função do “tradutor”. Também houve 29,9% das impressões em que o fluxo foi realizado entre a tela “Diciopédia” e “Termo Diciopédia”.

Os resultados apresentados nas seções 5.2.1 e 5.2.2, majoritariamente, validam o experimento realizado. Apesar de eles demonstrarem o engajamento dos usuários na realização do teste, o que pode ser indicativo de que a interface do aplicativo se apresenta de forma interessante ao consulente, eles não detalham as impressões dos usuários sobre cada tela e sobre a sua organização. Os mapas de calor baseados em toque (*touch heat maps*), que serão apresentados a seguir, trazem, por sua vez, informações sobre o uso do aplicativo no que concerne a esses aspectos.

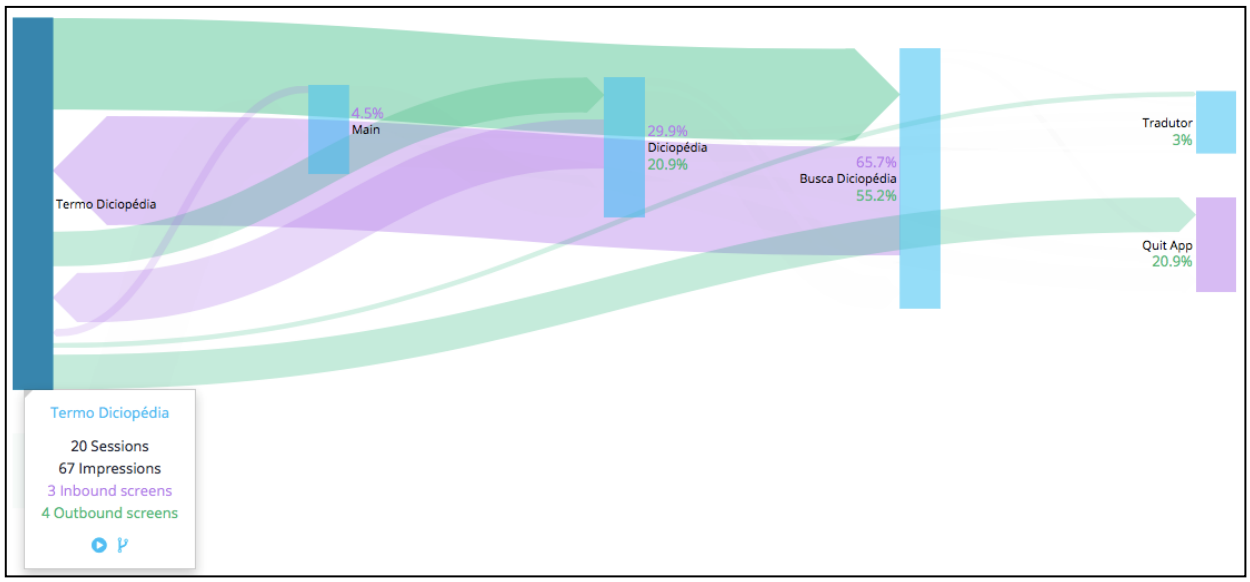


Figura 78: Resultado do fluxo do usuário para a tela de termos da Diciopédia.

Fonte: https://www.appsee.com/

5.2.3 Touch Heat Maps

O aspecto mais importante da análise visual do Appsee são os *heat maps* de toque compilados. A plataforma Appsee combina automaticamente todos os gestos de toque (toques, deslizamentos etc.) em cada tela do aplicativo para criar um mapa de calor de toque visualmente interpretável. Essa visualização nos permite entender rapidamente onde a maioria dos usuários está tocando em cada tela e o que está sendo ignorado. Também nos possibilita verificar se os usuários estão interagindo em uma determinada tela sem sucesso – casos conhecidos como **gestos não-responsivos**.

Ao observarmos as figuras apresentadas nesta seção, teremos um resumo de algumas informações que já foram mencionadas em outras seções, além do acréscimo de dados relevantes.

Na Figura 79, temos o apontamento da tela “Main” (principal). Nela, tivemos o registro de 30 sessões, correspondentes a 30 impressões, em uma duração média de 01:36 minuto, com o equivalente de 8,9% de interações do aplicativo. Foram 100% dos usuários que abriram o aplicativo, já que se trata da tela principal, sendo que, na sequência de uso, 56,7 % foram para a tela Diciopédia, 30% saíram do aplicativo, e 10% acessaram a tela “Termo Diciopédia”, como já se apontou anteriormente.

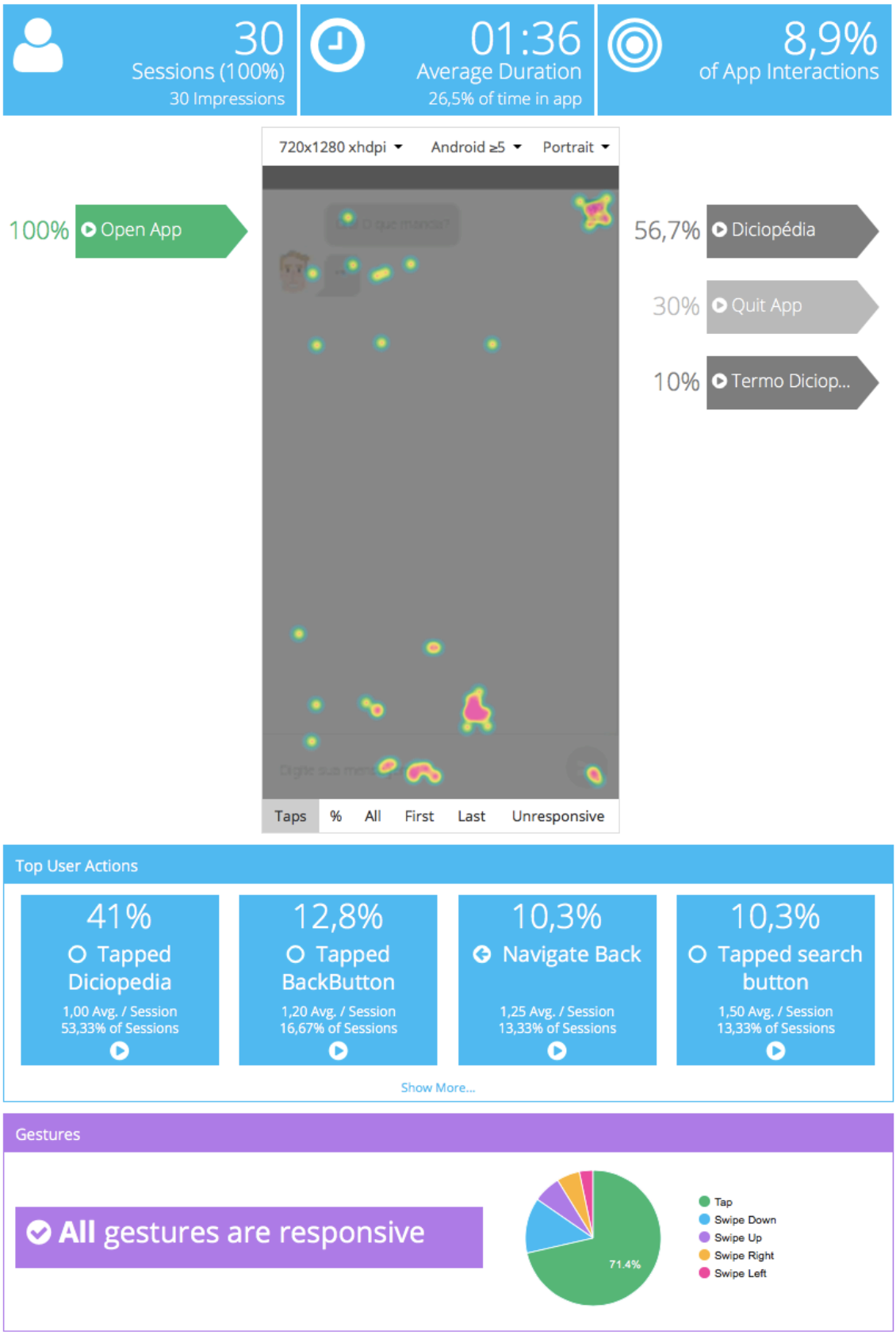


Figura 79: Touch heat map da tela main

Fonte: https://www.appsee.com/

Os novos dados trazidos por essa análise podem ser verificados em (i) plotagem das zonas de calor, onde temos uma marcação em pontos amarelos e rosa, que correspondem às áreas mais acessadas. Nesse caso, o principal tipo de interação registrado foi o toque, o qual concentra as quatro ações mais comuns na tela. Um aspecto a ser destacado é que os pontos de acesso à Diciopédia na interface, tanto o ícone “AZ” na parte superior, quanto o botão na parte inferior com o nome “Diciopédia” demonstraram ter sido claros para o usuário: trata-se dos locais com o maior índice de toques para acessar a função Diciopédia, mesmo que a informação sobre como acessá-la não tenha – propositalmente – sido oferecida nas instruções ao voluntário.

A segunda informação nova (ii) constatada na Figura 79 são as quantificações dos tipos de interação. Na parte em azul, temos as principais ações do usuário com os seguintes registros: 41% dos toques foram em Diciopédia; 12,8% clicaram no botão “voltar”; 10,8% clicaram em navegar de volta e 10,3% clicaram no botão de busca. Também temos a contagem de todos os dados responsivos dessa tela “Main” registrados por meio de um gráfico de pizza de acordo com os dados da legenda ao lado, sendo: 71,4 % tap (toque), o segundo percentual maior foi de *swipe down* (deslizar para baixo), depois de *swipe up* (deslizar para cima), na sequência *swipe right* (deslizar para a direita) e por último *swipe left* (deslizar para a esquerda).

Na sequência, temos os dados que foram registrados para a tela “Diciopédia” – Figura 80. Essa tela é a subsequente à tela “Main”. Foram registradas 20 sessões, com 41 impressões, com uma duração média de 02:01 minutos, correspondendo a 16,5% do uso de interação do aplicativo.

Dentre o percentual de acesso a cada uma das telas, novamente, retomaremos os seguintes dados: 41,5% tela “Main”, 34,1% tela “Termo Diciopedia”, “17% “Tradutor”, 48,8% “Termo Diciopédia”, 31,7 % “Busca Diciopédia” e 12,2 % saíram do aplicativo.

Quanto às plotagens da zona de calor, a maior concentração está no campo de busca e na porção da tela onde surge o teclado virtual para que as pessoas digitem. Também tivemos o registro de toques na parte inferior, nos balões de retorno à tela inicial. Mesmo não tendo sido orientados a usar a função “Tradutor”, que se encontra na parte superior da tela, ao lado da função Diciopédia, tivemos alguns incidentes, tal fato, se deve, possivelmente, por equívoco ou até mesmo por querer conhecer tal função. De toda forma, apontam para a necessidade de se repensar a inserção de um ponto de acesso a uma outra função do m.knob na mesma tela da Diciopédia.

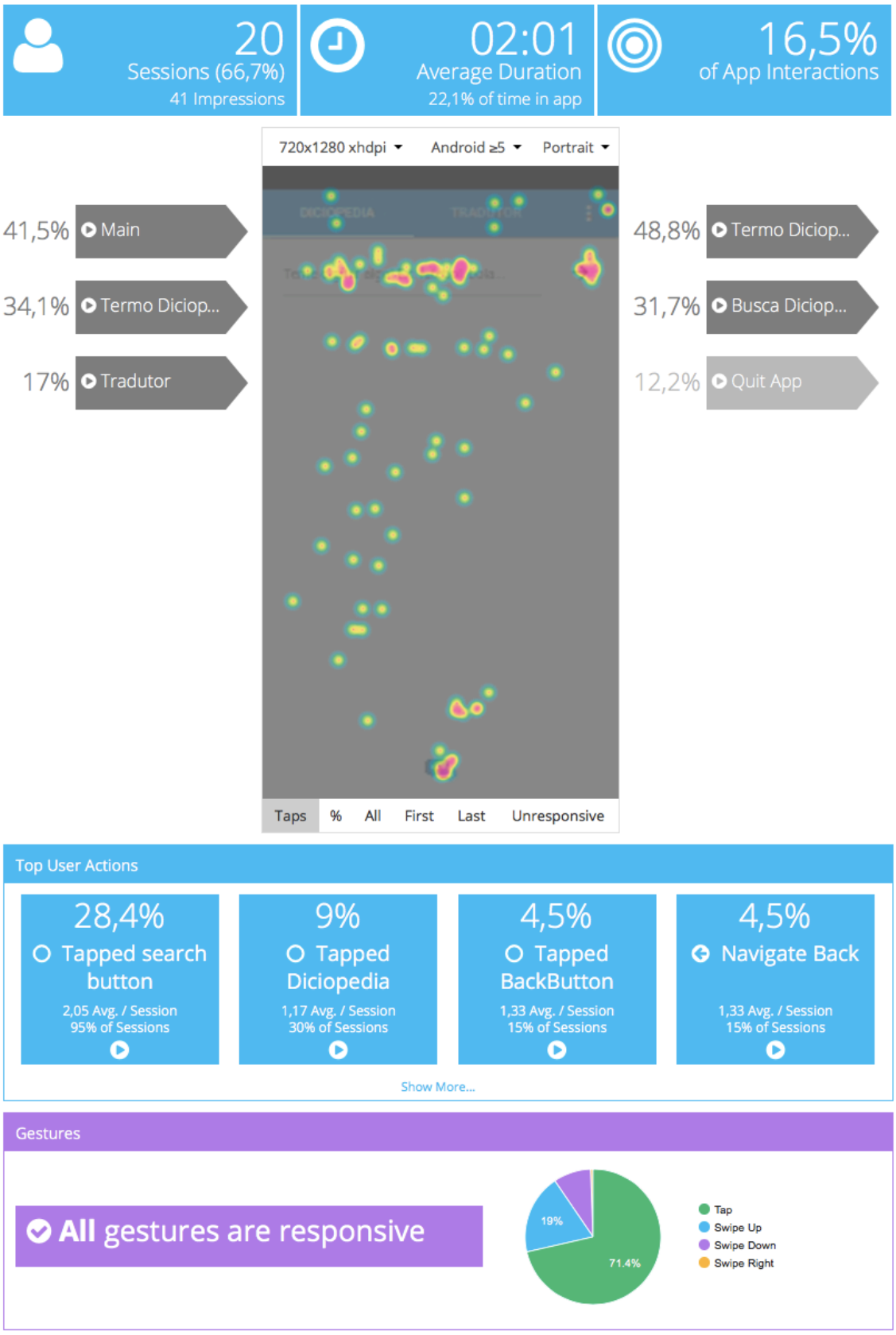


Figura 80: Touch heat map da tela de chegada da Diciopédia

Fonte: https://www.appsee.com/

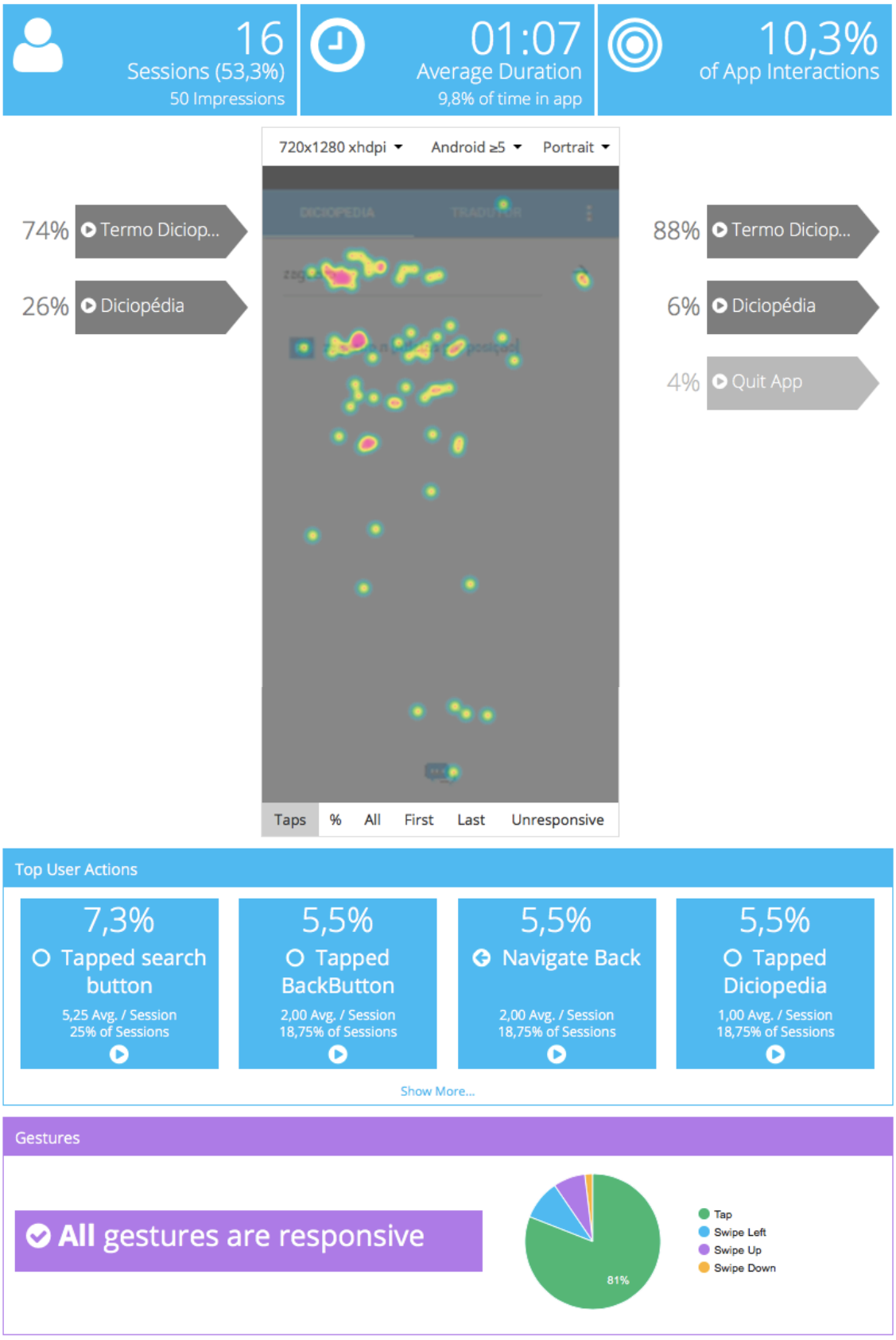


Figura 81: Touch heat map da tela de resultado de busca da Diciopédia

Fonte: https://www.appsee.com/

As principais ações dos usuários se concentraram no botão de busca sendo 28,4% dos toques, em seguida, tivemos 9% de toques em “Diciopédia”, depois, 4,5% de toques no botão que indica retorno, assim como 4,5 % de toques no botão navegar de volta.

No quesito de gestos, todos os dados responsivos dos usuários foram assim classificados de acordo com o gráfico e sua respectiva legenda que consta na última parte do relatório: 71,4 % de toques (tap), 19% *swipe up* (deslizar para cima) e em terceiro o *swipe down* (deslizar para baixo).

O próximo relatório (Figura 81) se refere à tela “Busca Diciopédia”, que é a aquela que traz os resultados de busca, assim como aquela que possibilita novas pesquisas sem ser necessário retornar à tela anterior. Nessa tela, obtivemos 16 sessões, que conferiram 50 impressões, com uma média de duração de 01:07 minuto, equivalente a 10,3% da interação do aplicativo.

Também foi registrado que 74% e 88% do fluxo de chegada e saída da tela ficaram direcionados para a tela “Termo Diciopédia”, reiterando que os voluntários realizaram sucessivas buscas na aplicação. Assim, a maior concentração dos *heat maps* se deu no local que digitamos uma nova busca, e nos resultados obtidos através dos verbetes. Lembramos que o teclado virtual também apresentou toques, conforme podemos averiguar na Figura 81.

Em relação às principais ações dos usuários, tivemos 7,3 % de toques no botão de busca e 5,5 % de toques para o botão de retorno, para navegar de volta e para Diciopédia.

No concernente aos gestos, ou seja, a todos os dados que são responsivos, novamente, o maior registro foi para o toque (*tap*) com 81%, e os demais foram *swipe left* (deslizar para esquerda), *swipe up* (deslizar para cima) e o menor índice foi *swipe down* (deslizar para baixo).

Na Figura 82, consta o *heat map* da tela “Termo Diciopédia”, ou seja, a tela em que encontramos o verbete para o termo que foi pesquisado. Obtivemos o registro de 20 sessões, com 67 impressões, com uma média de duração de 03:44 minutos, ocupando o maior ranqueamento de tempo e de interação do usuário com 61,3%.

Quanto ao fluxo, 65,7% e 55,2% correspondem à chegada a partir da tela “Busca Diciopédia” ou à saída para retornar a esta, respectivamente, reiterando, mais uma vez, as recorrentes interações de busca realizadas voluntariamente pelos participantes.

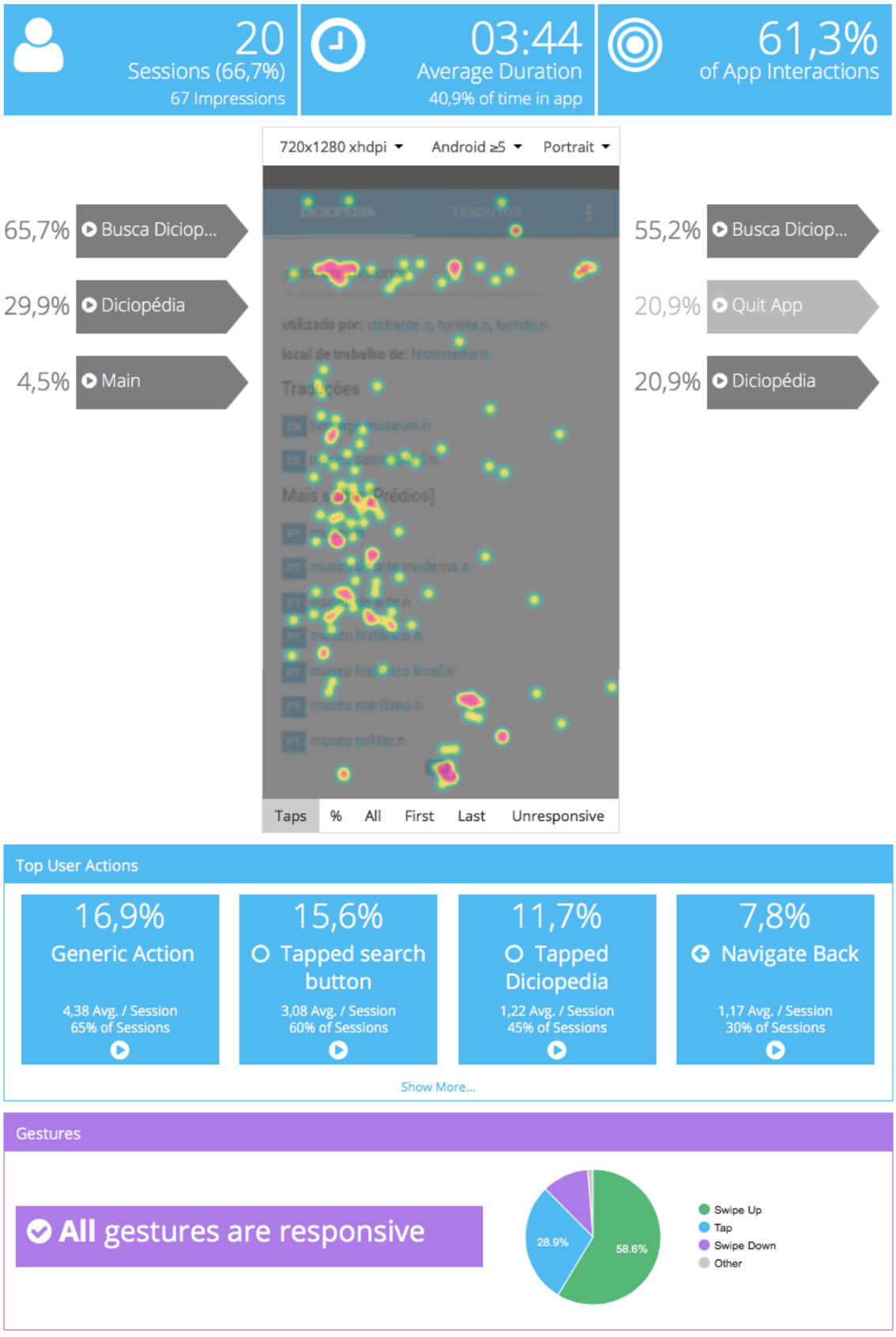


Figura 82: Touch heat map da tela de resultado de busca da Diciopédia

Fonte: https://www.appsee.com/

Nessa tela, podemos observar que os principais pontos focais de interação do usuário estão nas ferramentas de hipertexto, o que demonstra que tais soluções foram úteis e informativas, conforme os usuários responderam no questionário avaliativo da Diciopédia. Ressaltamos, aqui, que, conforme o Gráfico 12, 94,7% dos voluntários consideraram importantes as informações referentes às traduções; segundo o Gráfico 15, 81,6% dos consulentes afirmaram ter clicado nos hiperlinks, ou seja, nas informações que estão em cor azul na definição, corroborando também os dados do Gráfico 17, no qual 71,1% dos usuários consideraram muito relevantes as informações das estruturas qualia.

Os registros relacionados às principais ações dos usuários são os seguintes: 16,9% dizem respeito a ações genéricas, ou seja, a gestos que não resultaram em um clique, 15,6% clicaram no botão de busca, 11,7% clicaram em Diciopédia e 7,8% clicaram no botão para navegar de volta.

No que tange aos gestos, ou seja, a todos os dados que são responsivos, tivemos uma alteração em relação às telas analisadas anteriormente. Nessa tela “Termos da Diciopédia”, 58,6% foram *swipe up* (deslizar para cima), e apenas 28,9% correspondem a toques (*tap*), e em terceiro lugar tivemos *swipe down* (deslizar para baixo). Tal fato coincide com uma informação deixada nas observações e sugestões no final do questionário de avaliação da Diciopédia (Figura 71), na qual um dos usuários escreveu: “Em alguns casos, a relação das palavras sugeridas vai muito longe da busca inicial”.

Por fim, a Figura 83, que contrasta o primeiro gesto do usuário na tela – à esquerda – com o último gesto na mesma tela – à direita – reforça que a rolagem (*swipe*) se concentra na lista de “Mais sobre”, ou seja, de outros termos que são do mesmo frame, sendo, portanto, a primeira ação dos usuários para visualizar todas as informações disponíveis. Pelo fato de a rolagem, tanto para cima quanto para baixo, ter sido muito prevalente nessa tela, iremos propor algumas alterações de *design* que visem a minorar essa experiência negativa como o aplicativo, assim como em outros relacionados à interface da Diciopédia.

A partir de todos os dados e análises, constatamos muitos aspectos positivos em relação à função Diciopédia, objeto de análise desta tese, assim como verificamos que alguns detalhes ainda podem ser melhorados, de maneira a atender à demanda do consulente. Desse modo, no capítulo seguinte, propomos alguns ajustes tentando conciliar os aspectos relevantes trazidos pelas análises desse capítulo.

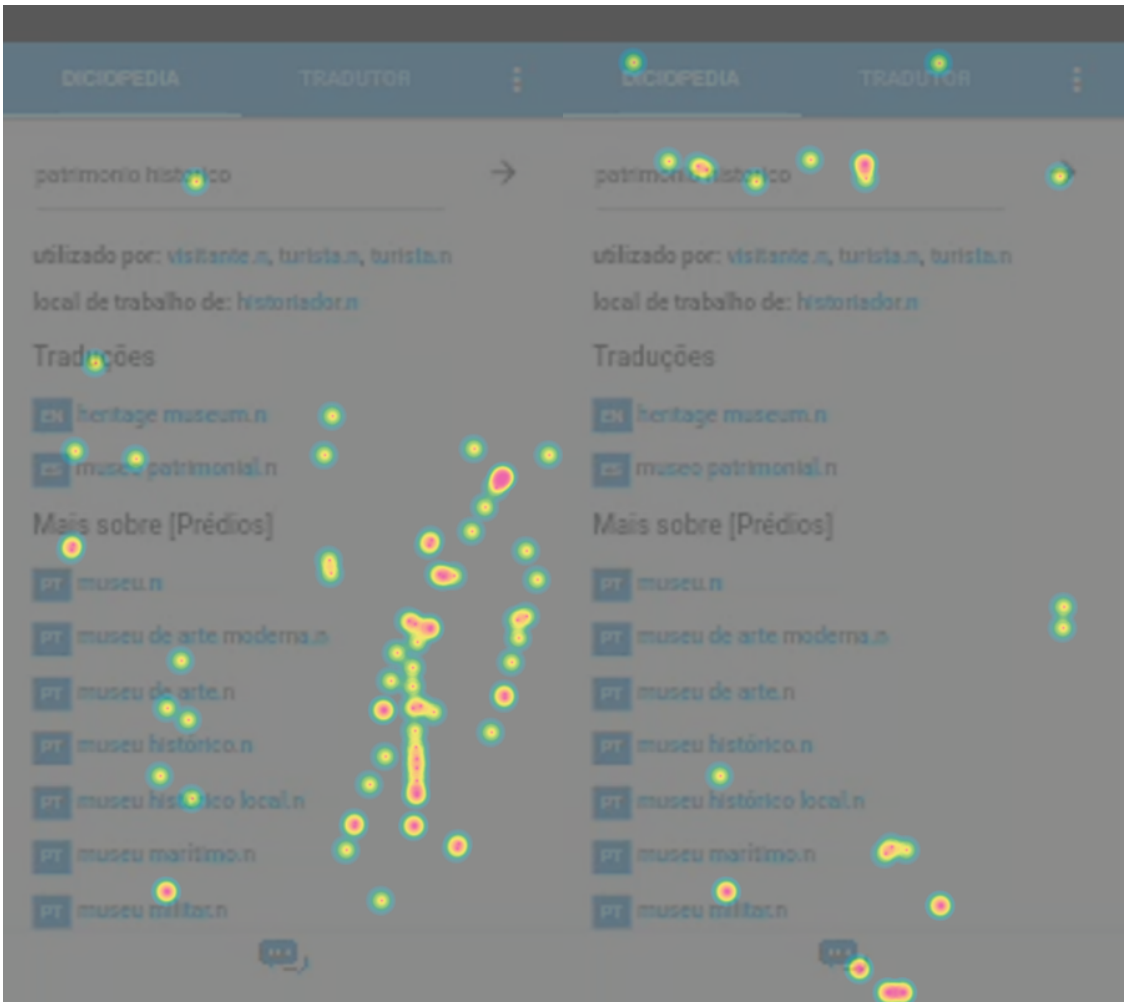


Figura 83: Touch heat map ampliado concernente à tela de Termos da Diciopédia

Fonte: https://www.appsee.com/

6 PROPOSTA DE NOVA INTERFACE PARA O VERBETE DA DICIOPÉDIA

Neste capítulo, nossa intenção é a de apresentar as alterações propostas para a interface da tela dos verbetes da Diciopédia, com base tanto nos dados do questionário quanto nos dados que nos foram fornecidos pela aplicação Appsee. Não teremos, neste momento, uma implementação de imediato na versão de produção do aplicativo, visto que novos ajustes requerem mudanças no todo e, para tal, temos uma demanda de tempo e de ajustes que envolvem toda a dinâmica do aplicativo.

A nova proposta de interface dos termos da Diciopédia, trará as informações mostradas na Figura 84.

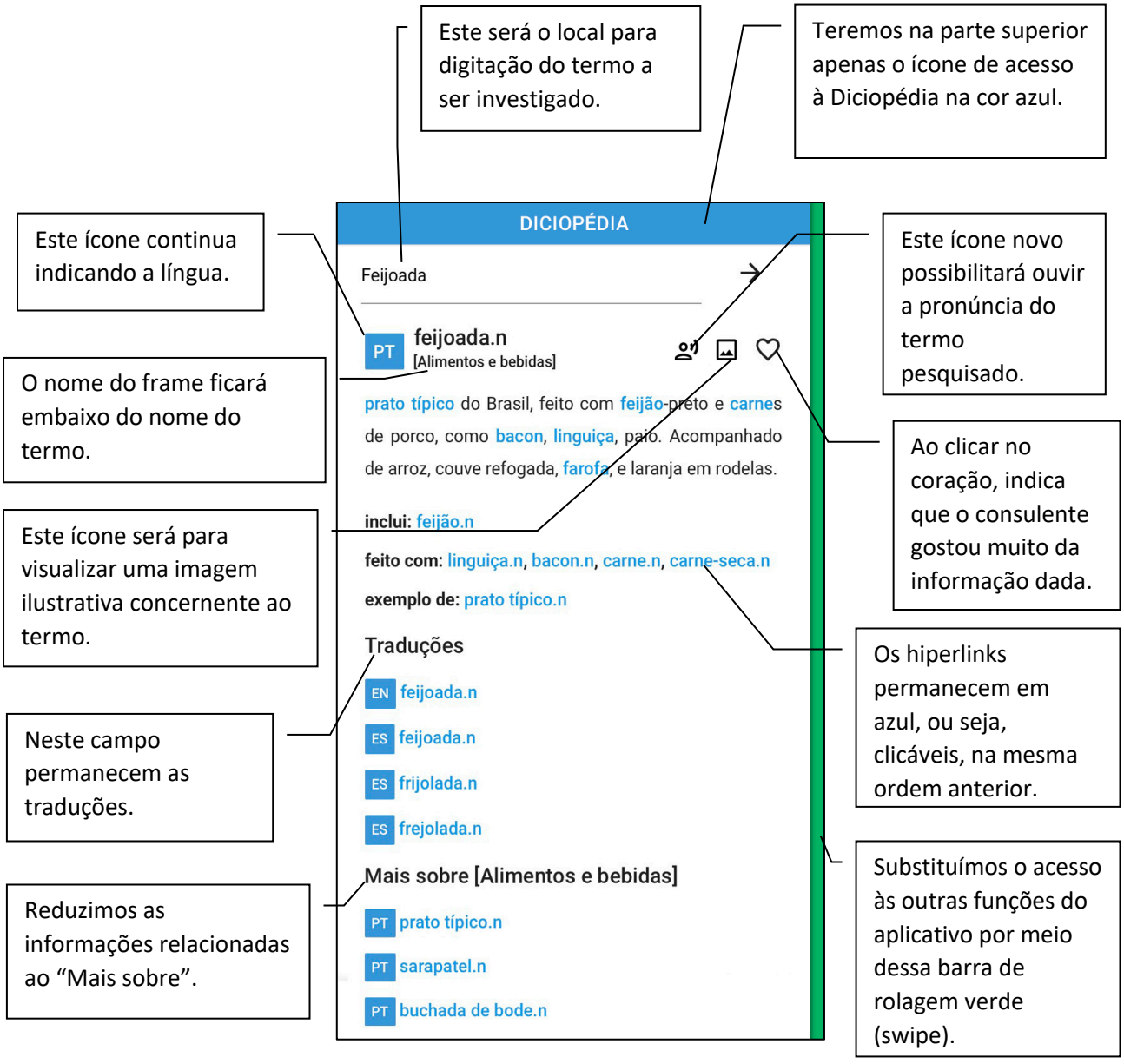


Figura 84: Proposta de design para a tela do verbete da Diciopédia

Mesmo sendo aparentemente simples, as alterações na tela do verbete da Diciopédia têm como finalidade facilitar e tentar atender as principais demandas dos consulentes, tentando explorar ao máximo não só as informações do nosso banco de dados, como também as tecnologias que estão disponíveis para os aparatos eletrônicos.

Primeiramente, teremos agora a individualização de acesso a cada uma das funções do aplicativo, via uma barra verde que está ao lado direito da tela. Por meio de rolagem (*swipe*), o consulente poderá acessar as outras funções, sendo que cada uma dessas telas terá uma cor que envolve as cores do aplicativo. Além de evitar problemas com o fluxo de navegação, essa solução também aumenta o espaço disponível na tela para a apresentação dos resultados: uma vez que a barra superior não é mais clicável (uma vez que o botão “Tradutor” foi removido), ela teve seu número de pixels reduzido.

O campo abaixo da barra azul na parte superior da tela, onde consta o nome “Diciopédia”, será o local onde o verbete poderá ser consultado. Para acessar novos termos, basta apagar o que foi digitado e escrever novos termos para consulta. Ao clicar na seta, obteremos os resultados que seguem abaixo do local da pesquisa.

Primeiramente, temos o ícone ISO 639-1 da língua em que o resultado está sendo fornecido. Depois, temos o nome do termo e a classe gramatical a que pertence (por exemplo, feijoada.n, sendo n, o indicativo para nome). Abaixo dessas informações, teremos o nome do frame evocado por esta unidade lexical, que anteriormente estava ao lado, porém, com o acréscimo dos ícones de escuta e de imagem, e a variação na extensão dos nomes dos frames, achamos melhor padronizar essa sequência.

Neste sentido, conforme algumas pessoas relataram no questionário, na Figura 71, consideramos ser importante poder ouvir a pronúncia do termo, independente da língua a ser escolhida, assim como poder acessar uma imagem que ilustre o termo que está sendo pesquisado. Evidentemente, grande parte das imagens será para mostrar principalmente os substantivos. Acreditamos que mesmo que demande algum tempo para conciliarmos tais informações, acreditamos que ela enriquecerá bastante os dados do aplicativo e da interação do usuário com ele. A imagem do coração ao lado direito já existia anteriormente e um clique nela favorita o termo buscado, armazenando-o no perfil do usuário.

Na sequência da Figura 84, temos a definição do termo e a possibilidade de acessar aos hiperlinks fornecidos tanto na definição quanto nas informações da estrutura qualia que estão abaixo.

Posteriormente, temos as traduções do verbete pesquisado tanto para o inglês quanto para o espanhol. Procuramos incluir, aqui, todas as variações possíveis que existem nessas línguas, mesmo que sejam diferentes, optamos por apresentar o maior número possível de termos.

Outra mudança muito importante e considerável, a partir dos dados dos *heat maps*, foi quanto às informações que apareciam referentes ao “Mais sobre”. Como em alguns casos temos uma listagem muito extensa de tais informações, procuramos reduzir a até no máximo cinco unidades lexicais mais importantes, ou seja, que apresentam o maior número de estruturas qualia compartilhadas. Assim, ao terminar a tela, pretendemos que o usuário não fique tendencioso a rolá-la (swipe) de maneira quase infinita, para verificar outras inúmeras informações.

Essas foram as sugestões que apresentamos para a equipe de desenvolvimento da FrameNet Brasil, de modo a tentar melhorar a função do m.knob avaliada nesta tese: a Diciopédia.

Evidentemente, com o avanço cada vez maior das ferramentas tecnológicas, do uso de aplicativos e de todo o processo tecnológico que nos cerca, sempre teremos ajustes a serem realizados para melhorarmos as aplicações, o que é natural e esperado na produção de e-dicionários.

No próximo capítulo, finalizamos esta tese retomando os principais aspectos que foram discutidos e que consideramos relevantes ao longo do que foi exposto até aqui.

7 CONCLUSÕES

O objetivo central desta tese foi o de verificar e quantificar os aspectos concernentes ao que é útil e relevante no dicionário eletrônico temático multilíngue Diciopédia, que foi desenvolvido a partir da Semântica de Frames. Para tal, percorremos o seguinte caminho nesta tese: abordamos os principais vieses teóricos e práticos da lexicografia, como, por exemplo, o seu aspecto interdisciplinar, e a importância da tomada de decisões que circundam o desenvolvimento de dicionários, tanto impressos quanto eletrônicos, sendo de linguagem especializada ou geral, monolíngues ou multilíngues, baseados em *corpora* ou não, entre tantos outros elementos que foram descritos no primeiro capítulo desta tese.

Para compreendermos os aspectos teóricos vinculados a prática lexicográfica, utilizamos o modelo desenvolvido por Atkins (2008, p.36), o qual foi aplicado contrastivamente a sete recursos terminológicos baseados em frames desenvolvidos anteriormente, no capítulo 2. Com essa revisão parametrizada de trabalhos anteriores, buscou-se identificar as características gerais do processo de tomada de decisões lexicográficas para e-dicionários terminológicos que possuem o mesmo cerne teórico que fundamentou a Diciopédia, considerando também que são eletrônicos e multilíngues.

Ainda no capítulo 2, apresentamos a estrutura de dados disponibilizada pela FrameNet Brasil, um recurso lexical baseado em frames e enriquecido com novas relações e estruturas qualia ternárias. Como a finalidade de todo produto lexicográfico é atender as principais necessidades e exigências do usuário, é preciso avaliar, respeitando-se o arcabouço teórico que sustenta cada recurso, quais dessas inúmeras informações a que temos acesso no banco de dados devem ser disponibilizadas para o consulente, sem sobrecarregá-lo com dados desnecessários, para que o uso seja eficiente, dinâmico e positivo.

Neste sentido, primeiro delimitamos a base teórica, já que, posteriormente, a partir dela, selecionamos as informações que consideramos mais relevantes e interessantes para o usuário, levando em conta o uso do dicionário. Tal aspecto responde ao problema de pesquisa que fora apresentado na introdução desta tese que era: “quais parâmetros metodológicos devem guiar o desenvolvimento de dicionários eletrônicos temáticos multilíngues baseados em recursos lexicais?”.

Considerando as análises dos dados do uso do aplicativo pelos voluntários, tivemos a validação da Diciopédia e de seus conteúdos, conforme havíamos projetado em nossa hipótese inicial de que a relevância da estrutura de dados que foi apresentada ao usuário correlaciona-se diretamente com a proeminência cognitiva dos frames, como modelos de representação da cognição linguística humana.

Ademais, os objetivos adjacentes foram igualmente alcançados com esta tese, uma vez que equalizaram-se os quantitativos de termos da Diciopédia, principalmente, no que tange ao domínio do Turismo, tendo em vista que a mesma metodologia que foi usada para o domínio do Turismo também está sendo realizada para o domínio dos Esportes, e, futuramente, poderá auxiliar até mesmo para outros domínios.

Também elaboramos, com base no arcabouço teórico da Semântica de Frames e da Lexicografia, a proposta de estruturação dos verbetes da Diciopédia, o que inclui a definição de termos, a tradução, assim como a inclusão das demais ferramentas que compõem tal função, como as de hipertexto e as colaborativas que estão presentes no aplicativo, o que caracteriza parte desse produto tecnológico inovador como o produto acadêmico de uma tese.[63] Tais aspectos também corroboraram de maneira significativa para o adensamento da base de dados da FrameNet, possibilitando um delineamento mais detalhado do léxico.

A partir da avaliação que realizamos através do questionário (Anexo 1) sobre a Diciopédia e das análises disponibilizadas pela plataforma Appsee, pudemos: (i) validar positivamente a função Diciopédia; (ii) constatar que as escolhas teóricas linguísticas foram relevantes e suficientes para a busca do usuário; (iii) verificar o que poderia ser melhorado no que tange tanto ao acréscimo quanto à redução de informações; (iv) propor alterações na interface da apresentação dos verbetes de acordo com as principais demandas dos usuários.

Além do produto tecnológico em si e desta tese, tivemos também um conjunto de outros produtos acadêmicos que foram produzidos como parte do desenvolvimento desses anos de pesquisa, tais como: participações em congressos nacionais e

[63] É importante frisarmos aqui que todos os trabalhos que são desenvolvidos na FrameNet Brasil não envolvem apenas um indivíduo. Para alcançarmos todos os dados que temos, dependemos de muitas pessoas envolvidas no projeto: pesquisadores, pós-graduandos, bolsistas e corpo técnico que, conjuntamente, implementam as discussões e propostas de cada tese e dissertação na forma de produtos tecnológicos.

internacionais na divulgação de nosso produto e do que vem sem produzido pelo laboratório da FrameNet Brasil, o que nos possibilitou ter as seguintes publicações:

- Capítulos de livros:

1. PERON-CORRÊA, S. R.; DINIZ, A.; TAVARES, T.; TORRENT, T. T. Contributions of frame semantics and construction grammar in the processes of machine translation. In: CUADRA, P. V.; REY, A. C.; GONZÁLEZ, P. C. (Orgs.). *Nuevas tendencias en traducción: Fraseología, Interpretación, TAV y sus didácticas*. Berlim: Peter Lang, 2018, v. 1, p. 531-552.
2. PERON-CORRÊA, S. R.; DINIZ, A.; LARA, M.; MATOS, E.; TORRENT, T. FrameNet-Based Automatic Suggestion of Translation Equivalents. In: SILVA, J.; RIBEIRO, R.; QUARESMA, P.; ADAMI, A.; BRANCO, A. (Orgs.). *Lecture Notes in Computer Science*. Berlim: Springer International Publishing, 2016, p. 347-352.
3. GONÇALVES, J.; PERON-CORRÊA, S. R.; SALOMÃO, M. M. M. Relación entre marcos y variación linguística: la traducción respaldada en la relación partitiva. In: HUERTA, P. M.; REY, A. C.; BLASCO, I. M.; BROTONS, L. N. (Orgs.). *Fraseología, variación y traducción*. Frankfurt: Peter Lang, 2016, v. 109, p. 205-222.
4. PERON-CORRÊA, S. R.; GONÇALVES, J.; TORRENT, T. T.; SALOMÃO, M. M. M. Contrate sintáctico-semántico de unidades léxicas del escenario del turismo en el portugués del Brasil y en el portugués europeo. In: HUERTA, P. M.; REY, A. C.; BLASCO, I. M.; BROTONS, L. N. (Orgs.). *Fraseología, variación y traducción*. Frankfurt: Peter Lang, 2016, v. 109, p. 255-266.

- Trabalhos completos publicados em anais de congressos:

1. DINIZ DA COSTA, A.; GAMONAL, M. A.; PAIVA, V. M. R. L.; MARÇÃO, N. D.; PERON-CORRÊA, S. R.; ALMEIDA, V. G.; MATOS, E. E.; TORRENT, T. T. FrameNet-Based Modeling of the Domains of Tourism and Sports for the Development of a Personal Travel Assistant Application. In: TORRENT, T. T.; BORIN, L.; BAKER, C. F. *Proceedings of the LREC 2018 Workshop International*

FrameNet Workshop 2018: Multilingual Framenets and Constructicons. Paris: European Language Resources Association, 2018, p. 6-12.

2. PERON-CORRÊA, S. R.; TORRENT, T. T. Constituição de Um Dicionário Eletrônico Trilíngue Fundado em Frames a partir da Extração Automática de Candidatos a Termos do Domínio do Turismo. In: Proceedings of STIL 2017 – XI Brazilian Symposium in Information and Human Language Technology and Collocated Events, 2017, p. 193-200.

Para concluirmos, relembramos que um dicionário não é apenas uma lista de palavras, mas, sim, um texto complexo que envolve uma multidimensionalidade de aspectos e de escolhas que precisam atender o consulente, possibilitando uma interação dinâmica. As propostas metodológicas que descrevemos sobre o desenvolvimento e o funcionamento da FrameNet Brasil possibilitam que os leitores desta tese compreendam os principais processos que são realizados para alcançarmos o desenvolvimento de uma aplicação, cujo produto final acessamos facilmente em um celular, mas que é composto por desenvolvimentos invisíveis primordiais para que os dados sejam os mais completos possíveis, além de serem também eficientes e interessantes ao consulente, que é o alvo principal.

Propor o desenvolvimento de um produto tecnológico em uma universidade pública não quer dizer que este produto será restrito apenas para os acadêmicos, pelo contrário, o intuito é mostrar as pesquisas que são potencialmente aplicáveis para o desenvolvimento de ferramentas tecnológicas que possam atender a sociedade e não apenas a universidade. Trabalhos como esses mostram a importância das pesquisas em universidades e, principalmente, da aplicação teórica de uma determinada área para fins de desenvolvimento de produtos que possam ser acessados por quem quer que seja.

Por último, temos visto que muitas aplicações têm sido desenvolvidas, contudo poucos são os trabalhos que se engajam em testes de usabilidade de tais aplicações. Neste sentido, consideramos que outra das contribuições desta tese foi a de propor um método replicável de avaliação da usabilidade de um produto lexicográfico, de modo a obter dados empíricos que permitam ao pesquisador entender os problemas relacionados à interface, interação e avaliação de e-dicionários.

REFERÊNCIAS BIBLIOGRÁFICAS

AGNOLONI, T. ET AL. A two–level knowledge approach to support multilingual legislative drafting. In: BREUKER, P. ET AL. (Ed.). *Law, ontologies and the semantic web.* Amsterdam: IOS Press, 2009.

ALAIN, B. *Le DiCoInfo. Del'idée à l'impression*. Rapport de stage. Département d'informatique et de recherche opérationnelle, Université de Montréal. 2006.

ALROOBAEA, R.; MAYHEW, P. J. How many participants are really enough for usability studies?. In: *2014 Science and Information Conference.* IEEE, 2014. p. 48-56.

ATKINS, B. T. S. Theoretical Lexicography and its Relation to Dictionary-making. In: FONTENELLE, T. (Ed.). *Practical Lexicography: a reader*. Oxford: Oxford University Press, 2008. p.31-50.

ATKINS, B. T. S.; LEVIN, B.; ZAMPOLLI, A. Computational approaches to the lexicon: Anoverview. In: ATKINS, B. T. S.; ZAMPOLLI, A. (Ed.). *Computational approaches to the lexicon*. Oxford University Press, 1994, p. 17-45.

BIDERMAN, J. C. O dicionário padrão da língua. In: *ALFA*. v. 28 (supl). São Paulo: UNESP, pp. 27-43, 1984.

BIDERMAN, J. C. A ciência da lexicologia. In: *ALFA*. v. 28 (supl). São Paulo: UNESP, pp.1-6, 1984.

BOAS, H. C. Frame semantics and translation. In: HOJO, A.; IBARRETXE-ANTUÑANO, I. (Ed.). *Cognitive linguistics and translation:* advances in some theorical models and applications. Berlin: De Gruyter, 2013, p.125-158.

BOAS, H. C. *Multillingual FrameNets in computational lexicography*: methods and applications. Berlin: De Gruyter, 2009.

BOAS, H. C. Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*, v. 18, n. 4, p. 445-478, 2005.

CABRÈ, M. T. Una nueva teoría de la terminología: de la denominación a la comunicación. In: CABRÉ, M. T. *La terminología*: representación y comunicación. Barcelona: Universitat Pompeu Fabra, 1999.

CALVI, M. V. Los géneros discursivos en la lengua del turismo: una propuesta de clasificación. *Ibérica***,** n. 19, p. 9-32, 2010. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=3185284>. Acessoem 26 out. 2017.

CASTILLO, R. A. *¿Cómo hacer un diccionario científico técnico?* Buenos Aires: Memphis, 1997.

CHISHMAN, R.; ET AL. FrameCup 2014: A Frame-based Domain-Specific Multilingual Electronic Dictionary. Comunicação oral na 12th International Cognitive Linguistics Conference – ICLC 12. Edmonton, Canada, 2013.

CRUSE, D. A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.

DA SILVA, B.; CHISHMAN, R. O papel dos frames na organização de dicionários online. *Calidoscópio*, v. 16, n. 3, 2018, p. 450-459.

DELL'ORLETTA, F.; MONTEMAGNI, S.; VENTURI, G. READ–IT: assessing readability of Italian texts with a view to text simplification. In: *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*. 2011. p. 73-83.

DINIZ DA COSTA, A. *A tradução por máquina enriquecida semanticamente com frames e estruturas qualia*. Qualificação (Progressão ao Doutorado) de Mestrado em Linguística. Juiz de Fora: Universidade Federal de Juiz de Fora, 2017.

DINIZ DA COSTA, A.; GAMONAL, M. A.; PAIVA, V. M. R. L.; MARÇÃO, N. D.; PERON-CORRÊA, S. R.; ALMEIDA, V. G.; MATOS, E. E.; TORRENT, T. T. FrameNet-Based Modeling of the Domains of Tourism and Sports for the Development of a Personal Travel Assistant Application. In: TORRENT, T. T.; BORIN, L.; BAKER, C. F. *Proceedings of the LREC 2018 Workshop International FrameNet Workshop 2018: Multilingual Framenets and Constructicons*. Paris: European Language Resources Association, 2018, p. 6-12.

DINIZ DA COSTA, A. *Frames e Qualia Ternários Aplicados aos Módulos de Pré e Pós-Processamento de um Pipeline de Tradução por Máquina Neural*. Tese de Doutorado em Linguística. Juiz de Fora: Universidade Federal de Juiz de Fora, em preparação.

DOLBEY, A.; ELLSWORTH, M.; SCHEFFCZYK, J. *BioFrameNet*: a domain–specific FrameNet extension with links to biomedical ontologies, O. BODENREIDER (Ed.). In: Proceedings of KR-MED, p. 87-94. 2006. Disponível em: <http://www.ceur-ws.org/Vol-222/krmed2006-p10.pdf>. Acesso em: 27 out. 2017.

DROUIN, P. Term extraction using non-technical corpora as a point of leverage. *Terminology,* v.9, n. 1, p. 99-115, 2003. Disponível em: <http://olst.ling.umontreal.ca/pdf/Terminology_2003.pdf>. Acesso em: 26 out. 2017.

ERK, K.; KOWALSKI, A.; PADO, S. *The salsa annotation tool-demo description*. In: Proceedings of the 6th lorraine-saarland workshop, p. 111-113, 2003. Disponível em: <https://pdfs.semanticscholar.org/e72c/cf884b7633afb32361504e269c4c537e36b8.pdf> . Acesso em: 27 out. 2017.

FABER, P. *et al.*Linkingimages and words: the description of specialized concepts. *International Journal of Lexicography*, v. 20, n. 1, p. 39-65, mar. 2007. Disponívelem: <https://doi.org/10.1093/ijl/ecl038>. Acesso em: 26 out. 2017.

FABER, P. *et al*. Process-oriented terminology management in the domain of coastal engineering. *Terminology*, v. 12, n. 2, p. 189-213, 2006. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.706.5899&rep=rep1&type=pdf>. Acesso em: 26 out. 2017.

FABER, P.; LEÓN ARAÚZ, P. Dinamismo en las bases de conocimiento terminológico. *Íkala, revista de lenguaje y cultura*, v. 15, n. 25, p. 75-100, 2010.

FABER, P.; C. MÁRQUEZ-LINARES; VEGA-EXPÓSITO, M. 2005. Framing terminology: a process-oriented approach. *META*, v. 50, n. 4. Disponível em: <https://www.erudit.org/fr/revues/meta/2005-v50-n4-meta1024/019916ar/>. Acesso em: 25 out. 2017.

FARIAS, E. M. P. A relação entre o léxico e o dicionário. *Revista de Letras*, v. 20, n. 1/2, p. 77-80, jan./dez. 1998. Disponível em: <http://repositorio.ufc.br/handle/riufc/16535 >. Acessoem: 26 out. 2017.

FELLBAUM, C. Large-scale lexicography in the digital age. *International Journal of Lexicography*, v. 27, n. 4, p. 378-395, dez. 2014.

FELLBAUM, C. *WordNet*. John Wiley & Sons, Inc., 1998.

FÉRNANDEZ, D. A. La lexicografiacomodisciplinalingüística. In: GUERRA, A. M. M. (Org.). *Lexicografiaespanõla.*Barcelona: Editorial Ariel, p. 31-52. 2003.

FILLMORE, C. C. J. Frame semantics.*Linguistics in the morning calm*, Hanshin Publishing Co., Seoul, South Korea, p. 111–137, 1982.

FILLMORE, C. J. Double-decker definitions: the role of frames in meaning explanations. *Sign Language Studies*, v. 3, n. 3, p. 263-295, 2003. Disponível em:<http://www.carstensinner.de/Lehre/spanisch/Material_Vidal_Diez/3_frames_y_explicaciones.pdf >. Acessoem: 27 out. 2017.

FILLMORE, C. J. *et al.* FrameNet in action: the case of attaching. *International Journal of Lexicography*, v. 16, n. 3, p. 297-332, 2003.

FILLMORE, C. J. Frame Semantics. In: BROWN, K. (Ed.). *Encyclopedia of language and linguistics.*Nueva York: Elsevier, 2006. p. 613-616.

FILLMORE, C. J. Frames and the semantics of understanding. Quaderni di Semantica, v. 6, n. 2, p. 222–254, 1985.

FILLMORE, C. J.; ATKINS, B. T. S. *FrameNet and lexicographic relevance*. In: Proceedings of the first International Conference on Language Resources and Evaluation (LREC 98), Granada, Espanha, p.417-423. Paris: European Language Resources Association. 1998.

FILLMORE, C. J.; ATKINS, B. T. S. Towards a frame-based lexicon: the semantics of risk and its neighbors. In: LEHRER, A.; KITTAY, E. F. (Ed.). *Frames, fields and contrasts*. Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates, 1992. p. 75-102. Disponívelem: <http://www.icsi.berkeley.edu/pubs/ai/towarda92.pdf >. Acessoem: 28 out. 2017.

FILLMORE, C.; BAKER, C.A frames approach to semantic analysis.The Oxford Handbook of Linguistic Analysis, n. 1, p. 313–340, 2010.

FILLMORE, C.; JOHNSON, C. R.; PETRUCK, M. R. Background to FrameNet.*International Journal of Lexicography*, v. 16, n. 3, p. 235–250, 2003. Disponívelem: <http://ijl.oxfordjournals.org/content/16/3/235.abstract>. Acesso em: 26 maio 2014.

FELLBAUM, C. Large-scalelexicography in the digital age. *InternationalJournalofLexicography*, v. 27, n. 4, p. 378-395, 2014. Disponível em: <https://academic.oup.com/ijl/article/27/4/378/933914> Acesso em: 27 de outubro de 2017.

FONTENELLE, T. (Ed.). *Practicallexicography: a reader*. Oxford University Press, 2008.p.1-15.

GAMONAL, M. A. Copa 2014 FrameNet Brasil: diretrizes para a constituição de um dicionário eletrônico trilíngue a partir da análise de frames da experiência turística. 2013. 146p. Dissertação de Mestrado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

_____. *Modelagem Linguístico-Computacional de Metonímias na Base de Conhecimento Multilíngue (m.knob) da FrameNet Brasil.* 175p. Tese de Doutorado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2017.

GANGEMI, A.; SAGRI, M.T; TISCORNIA, D.A constructive framework for legal ontologies. In: BENJAMINS, R. *et al* (Ed.). Law and the semantic web. Springer-Verlag Berlin, Heidelberg, 2005.

GROSS, G. “Comment ddcrireune langue de specialite?” Cahiers de lexicologie: Revue Internationale de Lexicologie et Lexicographie (80), 2002. pp. 179-200.

HOFFMANN, L. Conceitos básicos de uma lingüística das linguagens de especialidade. *Cadernos de tradução.* Porto Alegre, n. 17, p. 79-90, out./dez. 2004.

JOUSSE, A. L.; BOUVERET, M. Lexical functions to represent derivational relations in specialized dictionaries. *Terminology,* v. 9, n. 1, p. 71-98. 2003.

JOUSSE, A. L.; POLGUERE, A. *Le DiCo et sa version DiCouèbe. Document descriptif et manuel d’utilisation*. Version du rapport 1.0 – 19 avril 2005, Montréal: Observatoire de linguistique Sens-Texte (OLST).

KILGARRIFF, A. I don’tbelieve in wordsenses. *Computers and the Humanities*, v. 31, n. 2, p. 91-113, 1997.

KRALINGEN, R.; OSKAMP, E.; REURINGS, E. Norm frames in the representation of laws. In: SVENSSON, J.S.; WASSINK, J.G.J.; BUGGENHOUT, B. V. (Ed.). *Legal Knowledge Based Systems: Jurix '93: Intelligent Tools for Drafting Legislation, Computer-Supported Comparison of Law.*Lelystad: KoninklijkeVermande, 1993. p. 11-22. Disponível em: <http://www.egov.ufsc.br/portal/sites/default/files/anexos/2680-2674-1-PB.html >. Acesso em: 27 out. 2017.

KRIEGER, M.G. Tipologias de dicionários: registros de léxico, princípios e tecnologias. *Calidoscópio*, v. 4, n. 3, p. 141-147, 2006. Disponível em: <http://www.revistas.unisinos.br/index.php/calidoscopio/article/view/6000> Acesso em 28 de outubro de 2017.

KRIEGER, M. G.; FINATTO, M. J. B. Introdução à Terminologia: teoria e prática. EditoraContexto, 2004.

L'HOMME, M. C. Fonctions lexicales pour représenter les relations sémantiques entre termes. *Traitement automatique des langues* (*TAL*), v. 43, n. 1, p. 19-41, 2002.

L'HOMME, M. C. A methodology for describing collocations in a specialized dictionary. NIELSEN, S.; TARP, S. (Ed.). *Lexicography in the 21st Century In honour of Henning Bergenholtz*. Amsterdam/Philadelphia: John Benjamins. 2009.

L'HOMME, M. C. Définition du statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie,* v.73, n. 2, p. 61-84, 1998.

L'HOMME, M. C. Designing terminological dictionaries for learners based on lexical semantics: the representation of actants. In:FUERTES-OLIVERA, P. A.*Specialiseddictionaries for learners*. Berlin/New York: De Gruyter, 2010. cap. 8, p. 141-153.

L'HOMME, M. C. Designing terminological dictionaries for learners based on lexical semantics: the representation of actants. In: FUERTES-OLIVERA, P. (Ed.). *Specialised dictionaries for learners.* Berlin/New York: De Gruyter, 2010. p. 141-153.

L'HOMME, M. C. DiCoInfo: Dictionnaire fondamental de l'informatique et de l'Internet. *Consulté à l'adresse http://olst.ling. umontreal. ca/cgi-bin/dicoinfo/search. cgi*, 2009.

L'HOMME, M. C. *La terminologie:* principes et techniques. Montréal: Presses de l'Université de Montréal, 2004.

L'HOMME, M. C. Le DiCoInfo. méthodologie pour une nouvelle génération de dictionnaires spécialisés. *Traduire,* v.217, p. 78-103, 2008.

L'HOMME, M. C. Processing of terms in specializeddictionaries: new models and techniques. *Terminology*, v. 12, n. 2, 136 p. 2006.

L'HOMME, M. C. The processing of terms in dictionaries: new models and techniques: a state of the art. *Terminology:* International Journal of Theoretical and Applied Issues in Specialized Communication, v. 12, n. 2, p. 181-188, 2006.

L'HOMME, M.C. Why Lexical Semanticsis Important for E-Lexicography and WhyitisEqually Important to HideitsFormalRepresentationsfromUsers of Dictionaries. *InternationalJournalofLexicography*, v. 27, n. 4, p. 360-377, 2014. Disponível em: <https://academic.oup.com/ijl/article-abstract/27/4/360/935274>. Acessoem: 27 out. 2017.

L'HOMME, M. C.; B. Robichaud and C. Subirats.Discovering frames in specialized domains, In *Language Resources and Evaluation, LREC 2014*, Reykjavik, Iceland. 2014.

L'HOMME, M. C.; BODSON C.; VALENTE, R. S. Recherche terminographique semi-automatiséeenveilleterminologique: experimentation dans le domainemédical. *Terminologies nouvelles,* v. 20, p. 25-36, 1999.

L'HOMME, M. C.; CORMIER, M. C. Dictionaries and the digital revolution: A focus on users and lexical databases. *International Journal of Lexicography*, v. 27, n. 4, p. 331-340, 2014. Disponível em: <https://academic.oup.com/ijl/article/27/4/331/936324> Acesso em: 27 de outubro de 2017.

L'HOMME, M.C.; LANNEVILLE, M. È. DiCoEnviro. Dictionnaire fondamental de l'environnement. *Consulté à l'adresse http://olst. ling. umontreal. ca/cgibin/dicoenviro/search. cgi*, 2009.

L'HOMME, M. C.; ROBICHAUD, B. *Frames and terminology:* representingpredicativeterms in thefield of theenvironment. In: Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon, p. 186-197, 2014.Disponível em: <http://www.aclweb.org/anthology/W14-4723>. Acessoem: 27 out. 2017.

L'HOMME, M. C.; ROBICHAUD, B.; LEROYER, P. Encoding collocations in DiCoInfo: from formal to user-friendly representations. In: GRANGER, S.; PAQUOT, M. (Ed.). *ElectronicLexicography*, Oxford: Oxford UniversityPress. 2012.

L'HOMME, M. C.; SUBIRATS, C.; ROBICHAUD, B.A Proposal for combining" general" and specialized frames.In: *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*.2016. p. 156-165. Disponível em: <http://www.aclweb.org/anthology/W16-5321> Acesso em 28 de outubro de 2017.

LAKOFF, G. *Women, fire and dangerous things*. Chicago: University of Chicago Press, 1987.

LANDAU, M. Humanevolution as narrative: have heromyths and folk tales influenced our interpretations of the evolutionary past?. *American Scientist*, v. 72, n. 3, p. 262-268, 1984.

LARA, L. F..*Teoria del diccionario monolingüe*. Mexico, El Colégio de México: Centro de Estudios Lingüísticos y Literarios, 1996. p. 273.

LAVIOLA, A. B.; ET AL. The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil. In The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding Technical Report SS-17-02. Palo Alto, CA: AAAI Publications, v.17. p.193-196. 2017.

LE SERREC, A. *Étude sur l'équivalence de termes extraits automatiquement d'un corpus parallèle: contribution à l'extraction terminologique bilingue*. Mémoire de maîtrise de l'Université de Montréal. 2008.

LEMAY, C.; L'HOMME, M. C.; DROUIN, P. Two methods for extracting specific single-word terms from specialized corpora: experimentation and Evaluation. *International Journal of Corpus Linguistics,* v.10, n. 2, p. 227-255, 2005.Disponível em: <https://www.researchgate.net/publication/233631221_Two_methods_for_extracting_specific_single-word_terms_from_specialized_corpora_Experimentation_and_evaluation>. Acesso em: 27 out. 2017.

LEW, R.; DE SCHRYVER, G. M. Dictionaryusers in the digital revolution. *International Journal of Lexicography*, v. 27, n. 4, p. 341-359, 2014. Disponível em: <https://academic.oup.com/ijl/article/27/4/341/932743> Acesso em: 27 de outubro de 2017.

LÖNNEKER-RODMAN, B. Multilinguality and FrameNet. *International Computer Science Institute Technical Report*.Berkeley: ICSI, 2007. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.2092&rep=rep1&type=pdf>. Acessoem: 28 out. 2017.

LÓPEZ RODRÍGUEZ, C. I. *et al*. La terminología basada en marcos y su aplicación a las ciencias ambientales: los proyectos marcocosta y ecosistema. *Arena Romanistica*7, n. 10, p. 52-74, 2010. Disponível em: <https://www.researchgate.net/publication/235709205_La_Terminologia_basada_en_Marcos_y_su_aplicacion_a_las_ciencias_ambientales_Los_proyectos_MarcoCosta_y_Ecosistema>. Acesso em: 26 out. 2017.

MACIEL, A. M. B. Pertinência pragmática e nomenclatura de um dicionário terminológico. *Revista Internacional de Língua Portuguesa,* Lisboa, n. 15, p. 69-76, 1996.

MEL'CUK, I. *et al.Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques 1-IV*, Montréal : Les Presses de l'Université de Montréal. (1984-1999).

MEL'CUK, I. *et al*. Polguère. *Lexique actif du français*, Bruxelles:Duculot, 2007.

MEL'CUK, I.; A. Clas, A. et A. Polguère (1995). *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve (Belgique) : Duculot / Aupelf - UREF.

MEYER, I. Extracting knowledge-rich contexts for terminography: a conceptual and methodogical framework. In: BOURIGAULT, D.; JACQUEMIN, C.; L'HOMME, M. C. (Ed.). *Recent Advances in Computational Terminology.*Amsterdam, Filadelfia: John Benjamins, 2001. p. 279-302.

MOON, R. Sinclair, lexicography, and the Cobuild Project: the application of theory. *Internationaljournalof corpus linguistics*, v. 12, n. 2, p. 159-181, 2007.

OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (Org.). *As ciências do léxico:*lexicologia, lexicografia e terminologia. Campo Grande, MS: Ed. UFMS, 2001. 268 p.

PASTOR, V. & ALCINA, A. SearchTechniques in ElectronicDictionaries: a classification for translators. InternationalJournalofLexicography, 23 (3), p. 333357, 2010.

PAIVA, V. M. R. L. *Recomendação Automática de Atrações Turísticas a partir da Análise Semântica de Comentários de Usuários de Plataformas Colaborativas: uma aplicação da FrameNet Brasil.* Dissertação de Mestrado em Linguística. Juiz de Fora: Universidade Federal de Juiz de Fora, em preparação.

PERON-CORRÊA, S. R. *Copa 2014 FrameNet Brasil*: frames secundários em unidades lexicais evocadoras da experiência turística em português e em espanhol. 2014. 145 f.

Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

PERON-CORRÊA,S.R.*et al*. FrameNet-based automatic suggestion of translation equivalents. In: SILVA, J. *et al*. (Ed.). Computational Processing of the Portuguese Language.PROPOR 2016.Springer International Publishing, 2016. p. 347-352. (Lecture Notes in Computer Science, v. 9727).

PERON-CORRÊA,S.R. et al.*Contributions of Frame Semantics and Construction Grammar in the processes of Machine Translation.* In: Pino Valero Cuadra; AnalíaCuadrado Rey y Paola Carrión González (Ed.) Nuevastendenciasentraducción: Fraeología, Interpretación, TAV y susdidácticas. Peter Lang,2018,p. 531-552.

PETERS, W.; SAGRI, M. T.; TISCORNIA, D. The structuring of legal knowledge in lois. *Artificial Intelligence and Law,* v. 15, p. 117-135, 2007.

PETRUCK, M. Frame Semantics. En Östman, J. O.; Blommaert, J.; Bulcaen, C. (Ed.). *Handbook of pragmatics*. Ámsterdam/Philadelphia: Benjamins, 1996. p. 1-13.

PETRUCK, M. R. L. Framing motion in Hebrew and English.In: FAVRETTI, R. R. (Ed.). *Frames, corpora and knowledge representation.* Bologna: Bononia University Press, 2008.

POLGUERE, A. Collocations et fonctions lexicales: pour un modèle d'apprentissage. In : F. Grossmann et A. Tutin (éd.). *Les collocations. Analyse et traitement*, Coll. Travaux et recherches en linguistique appliquée, Paris: Éditions de Werelt, 2003. p. 117-142.

PRIETO- VELASCO, J. A. Hacia un modelo de uso de ilustraciones en el discurso científico-técnico. *Sendebar*, v. 18, p. 127-158, 2007.

PRIETO- VELASCO, J. A. Información gráfica y grados de especialidad en el discurso científico-técnico: un estudio de corpus. Tesis doctoral. Universidad de Granada. 2008.

PUSTEJOVSKY, J. *The generative lexicon*. Cambridge, USA: MIT Press, 1995.

PUSTEJOVSKY, J.; STUBBS, A. *Natural language annotation for machine learning*: a guide to corpus building for applications. Sebastopol, CA: O'Reilly,.p.342. 2012.

REY-DEBOVE, J. Léxico e dicionário .Trad. ClóvisBarleta de Morais. In: ALFA. V. 28 (supl). São Paulo: UNESP, p. 45-69.1984.

RIGUAL, C. C.; CALVI, M. V. Translation and lexicography: a necessary dialogue. In: *MonTi: Monografías de Traducción e Interpretación*, n. 6, p. 37-62, 2014. Disponível em: <http://www.raco.cat/index.php/MonTI/article/view/292089> .Acesso em 27 de outubro de 2017.

ROGERS, M. Multidimensionality in concepts systems: a bilingual textual perspective. *Terminology,* v. 10, n. 2, p. 215-240, 2004.

RUNDELL, M. Recent trends in english pedagogical lexicography1. In: FONTENELLE, Thierry (Ed.). *Practical lexicography: a reader*. Oxford UniversityPress, 2008.p.221 -243.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R. & SCHEFFCZYK, J. *FrameNet II: Extended Theory andPractice*. Berkeley, California: International Computer Science Institute, 2016.

SAGRI M–T. Tiscornia D.; Bertagna F.; Jur–WordNet, in Proceedings of the Second Global WordNet Conference, pp. 305–310, Brno, Czech Republic, January 20–23. 2004.

SARDINHA, T. B. Linguística de corpus. Rio de Janeiro: Manole, 2004.

SCHMIDT, T. The kicktionary: a multilingual resource of the language of football. In: REHM, G.; WITT, Andreas; LEMNITZER, L. (Ed.). *Data structures for linguistic resources and applications.*Tübingen: Gunter Narr, 2007. p. 189-196.

SCHMIDT, T. The Kicktionary: combining corpus linguistics and lexical semantics for a multilingual football dictionary. In: LAVRIC, E. *et al* (Ed.). The linguistic of football.Tübingen: Gunter Narr, 2008. p. 11-23.

SEELBACH, D. “Das kleinemultilingualeFußballlexikon”. In: Philologica et Linguistica. Historia, Pluralitas, Universitas. Festschrift fü r Helmut Humbachzum 80. Geburtstag am 4. Dezember 2001, edited by Bisang, Walter and Schmidt, Gabriela, Trier: Wissenschaftlicher Verlag. 2001, pp. 323-350.

SEELBACH, D. “Separable Partikelverben und Verbenmittypischen Adverbialen. Systematische Kontraste Deutsch-Französisch / Französisch-Deutsch”. In: Sprachtechnologiefü r die multilinguale Kommunikation. Beiträge der GLDV-Frühjahrstagung, edited by Seewald-Heeg, Uta, St. Augustin: Gardez.2003, pp. 103-115.

SINCLAIR, J. M.; et al. Progress in English computational lexicography. *World Englishes*, v. 9, n. 1, p. 21-36, 1990.

SIGILIANO, N. S.; TORRENT, T. T. FrameNet annotation as a means to identify genre-relevant linguistic structures. *Scriptum*, v. 3, no prelo.

STORRER, A. “Digitale Wörterbücherals Hypertexte: ZurNutzung des H ypertextkonzepts in der Lexikographie”. In: Chancen und Perspektivencomputergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikationdigitaler Wörterbücher, edited by Lemberg, Ingrid; Schröder, Bernhard and Storrer, Angelika, Tübingen: Niemeyer, 2001. pp. 88-104.

TEMMERMAN, R. 1997. Questioning the univocity ideal: the difference between sociocognitive terminology and traditional terminology.*Hermes, Journal of Linguistics*, n. 18: 51-91. Disponívelem: <https://www.researchgate.net/publication/313590021_Questioning_the_univocity_ideal_The_difference_between_socio-cognitive_Terminology_and_traditional_Terminology>. Acessoem: 26 out. 2017.

TEMMERMAN, R. *Towards new ways of terminology description:* the sociocognitive approach. Amsterdam/Philadelphia: John Benjamins, 2000.

TORRENT, T. T.; ELLSWORTH, M. Behind the Labels: criteria for defining analytical categories in FrameNet Brasil. Veredas, 17 (1), p. 44-65, 2013.

TORRENT, T. T., ET AL. (2014a). Copa 2014 FrameNet Brasil: a framebased trilingual electronic dictionary for the Football World Cup. In COLING (Demos) (pp. 1014).

TORRENT, T. T., ET AL. (2014b). Multilingual lexicographic annotation for domainspecific electronic dictionaries: The Copa 2014 FrameNet Brasil project. Constructions and Frames, 6(1), 73-91.

TORRENT, T. T., ET AL. (no prelo). A flexible tool for an enriched FrameNet: the FrameNet Brasil Webtool. 2017.

VAN KRALINGEN, R.; OSKAMP, E.; REURINGS, E. Norm frames in therepresentation of laws. *Legal Knowledge Based Systems: Jurix*, v. 93, 1993.

VENTURI, G. *et al*. *Towards a FrameNet resource for the legal domain.* In: Proceedings of the IIIth Workshop on legal ontologies and artificial intelligence techniques (LOAIT ‟09). Barcelona. 2009.

VENTURI, G. *Parsing legal texts:* a contrastive study with a view to knowledge management applications. In: Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC 2008), Workshop Semantic Processing of Legal Texts, Marrakech, Morocco, May 26-1 June, 2008. 1 CD-ROM.

VISSER, P. R. S. *Knowledge specification for multiple legal tasks:* a case study of the interaction problem in the legal domain. The Hague, The Netherlands: Kluwer Law International, 1995. 250p. (Computer/Law Series, 17).

ZAMPOLLI, A. Introduction. In: ATKINS, Beryl T. Sue; ZAMPOLLI, Antonio (Ed.). Computational approaches tothelexicon. Oxford University Press, p. 3-16, 1994.

Dicionários

HOUAISS, Antônio; VILLAR, Mauro de Salles. Dicionário Houaiss da língua portuguesa. Rio de Janeiro: Objetiva, 2009.

http://michaelis.uol.com.br/moderno-portugues/como-consultar/organizacao-do-verbete/

Recursos computacionais

Dicionário da Copa 2014. Disponível em: <http://dicionariodacopa.com.br/.>Acesso em: 29 de out. de 2017.

Dicionário Field. Disponível em: <http://dicionariofield.com.br/>Acesso em: 29 de out. de 2017.

DicoEnviro. Disponível em: <http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi>Acesso em: 29 de out. de 2017.

DicoInfo. Disponível em: <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>Acesso em: 29 de out. de 2017.

EcoLexicon. Disponível em: <http://manila.ugr.es/visual/>Acesso em: 29 de out. de 2017.

FRAMENET, 2017. Disponível em: <https://framenet.icsi.berkeley.edu/fndrupal/> Acesso em: 29 de out. de 2017.

WEBTOOL. 2017. Disponível em: <http://webtool.framenetbr.ufjf.br/> . Acesso em: 29 de out. de 2017.

Kicktionary. Disponível em: <http://www.kicktionary.de/index.html>Acesso em: 29 de out. de 2017.

MKNOB, 2017. Disponível em: <http://mknob.com/> Acesso em: 29 out. 2017.

SKETCH ENGINE. 2017. Disponível em: <https://the.sketchengine.co.uk>. Acesso em: 29 out. 2017.

ANEXO I

Questionário de impressões de uso da função Diciopedia do Aplicativo m.knob

Prezado Voluntário,

Obrigada por se dispor a responder essa rápida pesquisa acerca das suas impressões de uso da função Diciopedia do Aplicativo m.knob.

O questionário está dividido em 4 partes, quais sejam:
1. Dados do Participante e Termo de Consentimento;
2. Sobre a Experiência do Participante com Dicionários e Idiomas;
3. Sobre a Experiência do Participante com a Diciopedia;
4. Sobre os Verbetes da Diciopedia.

Para responder a seção 2, você será convidado a utilizar o aplicativo m.knob. Siga atentamente as instruções que lhe forem dadas.

Mais uma vez, obrigada pela colaboração!

Simone Rodrigues Peron-Corrêa
Doutoranda em Linguística – Universidade Federal de Juiz de Fora

*Obrigatório

1. **Endereço de e-mail** *

1. Dados do participante e termo de consentimento

2. **Nome** *

3. **Documento de Identidade** *

4. **Endereço Completo** *

Termo de Consentimento

O Sr. (a) está sendo convidado (a) como voluntário (a) a participar da avaliação de um dicionário eletrônico temático. Para participar deste estudo o Sr (a) não terá nenhum custo, nem receberá qualquer vantagem financeira. Terá o esclarecimento sobre o estudo em qualquer aspecto que

desejar e estará livre para participar ou recusar-se a participar. Poderá retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que é atendido pela pesquisadora, que tratará a sua identidade com padrões profissionais de sigilo.
A identidade do sr.(a) será mantida em sigilo na divulgação dos resultados da pesquisa e, por se tratar de pesquisa linguística, os dados fornecidos por você não serão objeto de nenhum tipo de avaliação social, e, por isso, sua participação não implica riscos nem benefícios pessoais.
Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não serão liberados sem a sua permissão.

5. **Você está de acordo com o termo de consentimento?**
   Marcar apenas uma oval.

   - ◯ Estou de acordo
   - ◯ Não estou de acordo *Ir para "Obrigada pela sua participação!".*

1. Sobre sua experiência com dicionários e idiomas

As perguntas a seguir têm por objetivo colher informações sobre a sua experiência prévia com dicionários e idiomas.

6. **Você costuma usar dicionários?** *
   Marcar apenas uma oval.

   - ◯ Sim
   - ◯ Não

7. **Se respondeu "sim", qual você acessa mais?**
   Marcar apenas uma oval.

   - ◯ Eletrônico
   - ◯ Papel

8. **Qual seu nível de proficiência em inglês?** *
   Marcar apenas uma oval.

	1	2	3	4	5	
Não falo inglês	◯	◯	◯	◯	◯	Sou fluente em inglês

9. **Qual seu nível de proficiência em espanhol?** *
   Marcar apenas uma oval.

	1	2	3	4	5	
Não falo espanhol	◯	◯	◯	◯	◯	Sou fluente em espanhol

ATENÇÃO: Leia atentamente as instruções a seguir antes de continuar

Para responder as perguntas da próxima seção, será necessário que use a Diciopedia. Para tanto, você deve:
1. Clicar no ícone do m.knob na tela do celular que lhe será fornecido pela pesquisadora;
2. Acessar a função Diciopedia: dicionário temático dos campos do Turismo e dos Esportes;
3. Buscar as palavras de sua preferência e interagir com a aplicação da forma que achar mais

conveniente;
4. Retornar à tela inicial do m.knob quando terminar, clicando nos balões localizados na parte inferior da tela da Diciopedia.

ATENÇÃO: a função TRADUTOR não está sob avaliação e, portanto, não deve ser utilizada.

2. Sobre a sua experiência com a Diciopedia

As perguntas a seguir têm por objetivo avaliar a sua experiência de uso com a Diciopedia.

10. **Considerando os campos cobertos pelo mknob (Turismo e Esportes Olímpicos), no geral, como você avalia os resultados das buscas pelas palavras desejadas?** *
    Marcar apenas uma oval.

    - ◯ Resultado encontrado
    - ◯ Não houve nenhum resultado
    - ◯ O resultado apresentado não era adequado à busca que tinha em mente

11. **Caso não tenha obtido algum resultado desejado, qual(is) palavra(s) foi(ram) buscada(s)?**

    _____

12. **Você chegou a sugerir este(s) termo(s) no aplicativo?**
    Marcar apenas uma oval.

    - ◯ Sim
    - ◯ Não

13. **Quão fácil foi a utilização da Diciopedia?** *
    Marcar apenas uma oval.

	1	2	3	4	5	
Muito difícil	◯	◯	◯	◯	◯	Muito fácil

14. **Considerando-se que a Diciopedia é uma das funcionalidades de um app voltado para turistas, como você avalia a utilidade dessa funcionalidade para o público-alvo do app?** *
    Marcar apenas uma oval.

	1	2	3	4	5	
Totalmente inútil	◯	◯	◯	◯	◯	Muito útil

15. **Você considera as traduções uma informação importante na Diciopedia?** *
    Marcar apenas uma oval.

    - ◯ Sim
    - ◯ Não
    - ◯ Não reparei na existência de traduções

16. **Como você avalia as traduções apresentadas pela Diciopedia?**
Marcar apenas uma oval.

	1	2	3	4	5	
Muito ruins	⬭	⬭	⬭	⬭	⬭	Muito boas

3. Sobre os verbetes da Diciopedia

As perguntas a seguir têm por objetivo avaliar sua percepção sobre as informações disponibilizadas nos verbetes da Diciopedia.

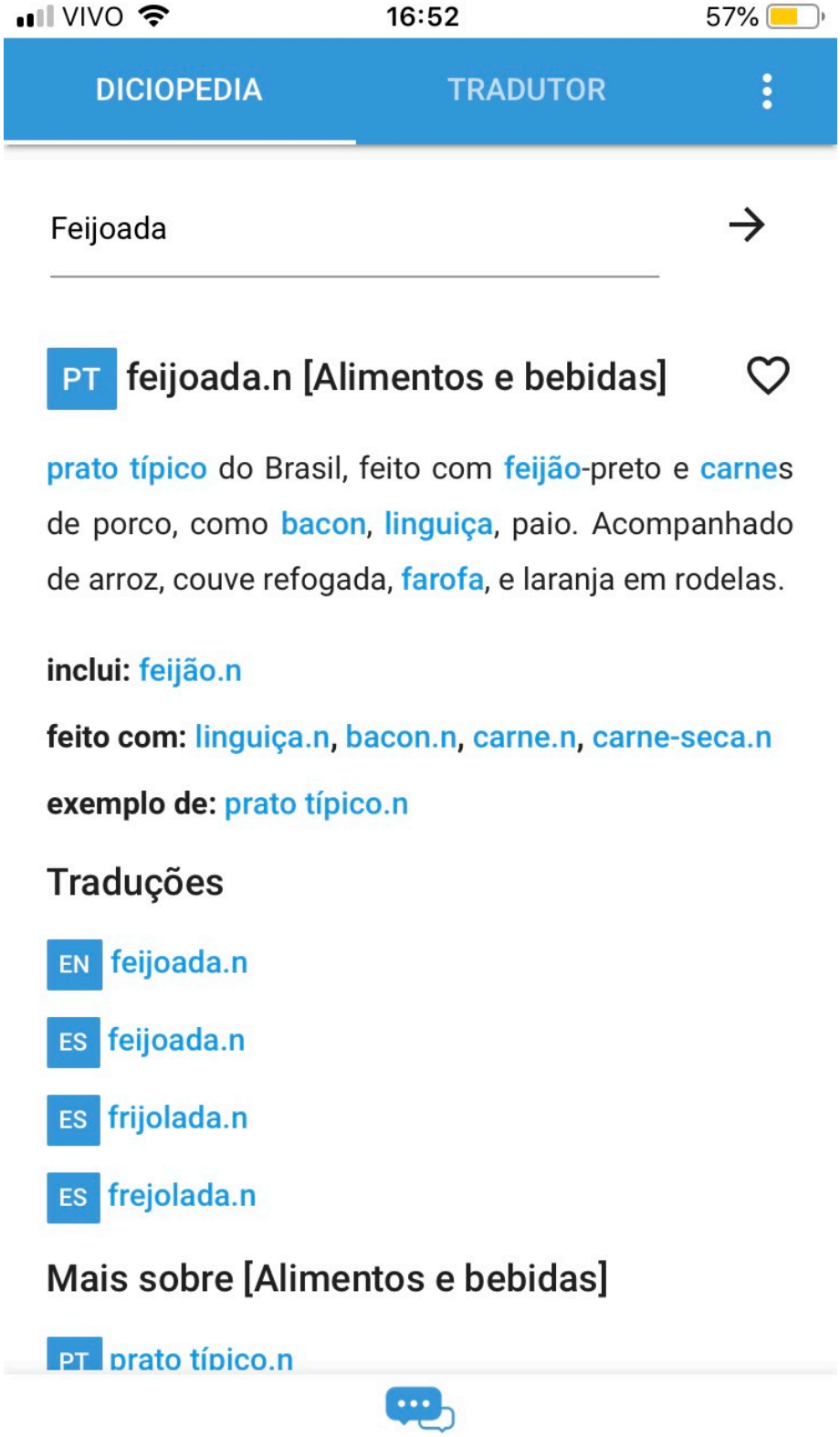
17. **Qual a relevância da informação apresentada entre colchetes ao lado da lista de resultados de uma pesquisa?** *



Marcar apenas uma oval.

	1	2	3	4	5	
Irrelevante	◯	◯	◯	◯	◯	Muito relevante

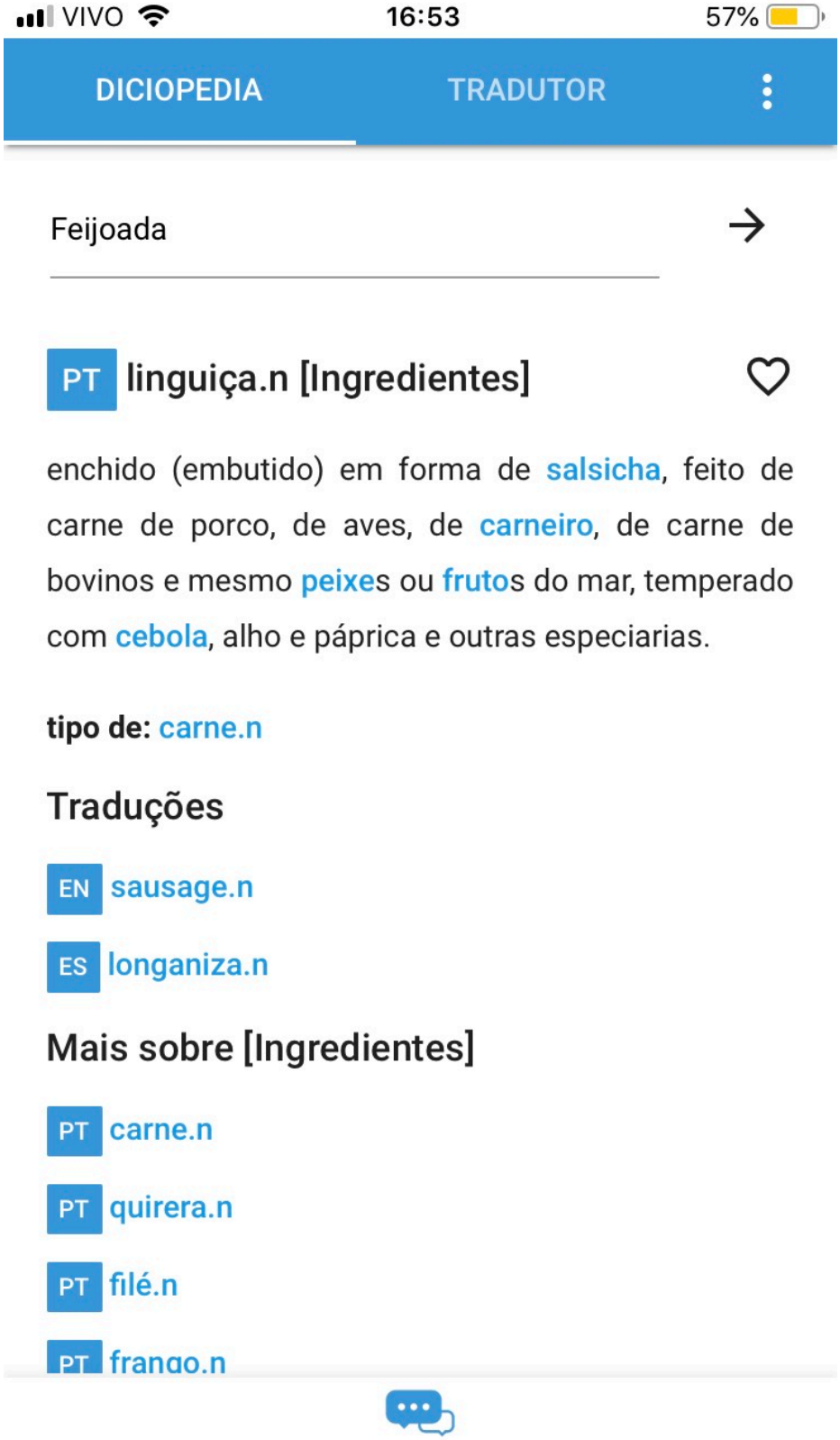
18. **Considerando a tela que apresenta o resultado da busca, você chegou a clicar nas palavras que estavam escritas em cor azul em meio à definição do termo buscado?** *



Marcar apenas uma oval.

- ◯ Sim
- ◯ Não

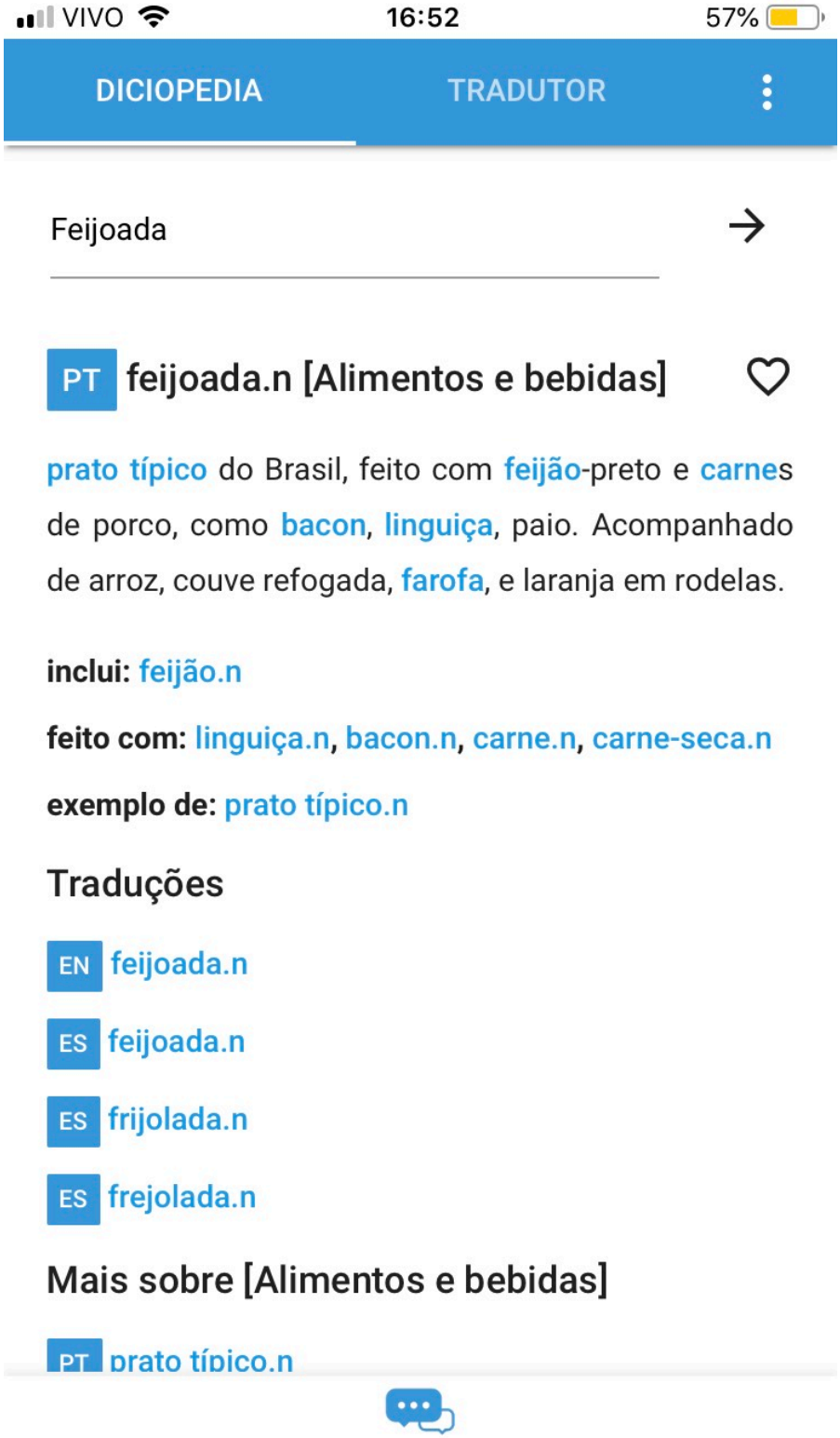
19. **Como você avalia a relevância das informações que foram acessadas ao se clicar nas palavras em cor azul, conforme o exemplo abaixo?** *



Marcar apenas uma oval.

	1	2	3	4	5	
Totalmente irrelevantes	◯	◯	◯	◯	◯	Muito relevantes

20. **Como você avalia a relevância das informações que são apresentadas abaixo da definição da palavra, tais como, “inclui”, “feito com”, “exemplo de” entre outras, conforme o exemplo abaixo** *



Marcar apenas uma oval.

	1	2	3	4	5	
Totalmente irrelevantes	◯	◯	◯	◯	◯	Muito relevantes

21. **A última informação trazida pela Diciopedia é uma lista de palavras que estão relacionadas à palavra buscada, na opção "Mais sobre". Você chegou a consultar alguma dessas palavras?** *

Marcar apenas uma oval.

- ◯ Sim
- ◯ Não

22. **Essas palavras listadas ampliaram o seu leque de busca?** *

Marcar apenas uma oval.

- ◯ Sim
- ◯ Não

23. **Você considera que essa lista de palavras pode auxiliar na definição da palavra relacionando-a em uma rede de significados?** *

Marcar apenas uma oval.

- ◯ Sim
- ◯ Não

24. **Como você avalia as informações dadas pela Diciopedia?**

Marcar apenas uma oval.

	1	2	3	4	5	
Incompletas	◯	◯	◯	◯	◯	Muito completas

25. **Se você considera incompletas, que tipo de informação ou dado incluiria?**

_____

_____

_____

_____

_____

Obrigada pela sua participação!

Powered by
Google Forms