



**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

**ACESSO E COMPREENSÃO DE CONTEÚDOS EM PORTUGUÊS POR
ESTRANGEIROS EM BIBLIOTECAS DIGITAIS CIENTÍFICAS: UMA PROPOSTA
METODOLÓGICA E SUA IMPLEMENTAÇÃO**

Francisco Cláudio Sampaio de Menezes

Brasília, DF
2017

Francisco Cláudio Sampaio de Menezes

ACESSO E COMPREENSÃO DE CONTEÚDOS EM PORTUGUÊS POR
ESTRANGEIROS EM BIBLIOTECAS DIGITAIS CIENTÍFICAS: UMA PROPOSTA
METODOLÓGICA E SUA IMPLEMENTAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília para obtenção do título de Doutor em Ciência da Informação, na área de concentração em gestão da informação.

Orientadora: Prof.^a Dr.^a Dulce Maria Baptista

Brasília, DF
2017

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SF819a Sampaio de Menezes, Francisco Cláudio
ACESSO E COMPREENSÃO DE CONTEÚDOS EM PORTUGUÊS POR
ESTRANGEIROS EM BIBLIOTECAS DIGITAIS CIENTÍFICAS: UMA
PROPOSTA METODOLÓGICA E SUA IMPLEMENTAÇÃO / Francisco
Cláudio Sampaio de Menezes; orientador Dulce Maria
Baptista. -- Brasília, 2017.
197 p.

Tese (Doutorado - Doutorado em Ciência Política) --
Universidade de Brasília, 2017.

1. Alinhamento Sentencial. 2. Biblioteca Digital. 3.
Multilinguismo. 4. Sumarização Automática. 5. Tradução por
Máquina. I. Baptista, Dulce Maria, orient. II. Título.

FOLHA DE APROVAÇÃO

Título: "Acesso e Compreensão de Conteúdos em Português por Estrangeiros em Bibliotecas Digitais Científicas: Uma Proposta Metodológica e Sua Implementação"

Autor (a): Francisco Cláudio Sampaio de Menezes

Área de concentração: Gestão da Informação

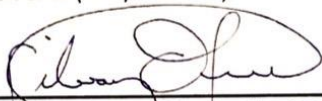
Linha de pesquisa: Organização da Informação

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade de Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **Doutor** em Ciência da Informação.

Tese aprovada em: 29 de novembro de 2017.



Profª Drª Dulce Maria Baptista
Presidente (UnB/PPGCINF)



Profª Drª Gilvan Müller de Oliveira
Membro Externo (UFSC)



Profª Drª Sabine Gorovitz
Membro Externo (INT/UNB)



Profª Drª Murilo Bastos da Cunha
Membro Interno (UnB/PPGCINF)

Profª Drª Rogério Henrique de Araújo Júnior
Suplente (UnB/PPGCINF)

À Marilde, companheira de todas as horas ao longo da vida

AGRADECIMENTOS

Fruto de uma longa trajetória, uma tese de doutorado contém e representa uma importante etapa da minha vida acadêmica e profissional resultante de muito esforço e de muitas ajudas.

No plano institucional, primeiramente, quero registrar meu apreço à Universidade de Brasília (UnB) que – através do Departamento de Línguas Estrangeiras e Tradução (LET), do Instituto de Letras (IL) – me propiciou o apoio necessário ao desenvolvimento desta pesquisa.

Aos professores da Faculdade de Ciência da Informação da UnB, cuja contribuição intelectual inegável muito me ajudou à realização deste trabalho, presto aqui meu preito de reconhecimento. Nesse particular, não posso deixar de enaltecer o trabalho da minha orientadora, Prof.^a Dulce Maria Baptista, que, em todos os momentos do desenvolvimento da tese, me apoiou incondicionalmente tanto no plano intelectual e acadêmico como no plano do bom relacionamento cotidiano entre orientando e orientador. Esse estímulo diuturno e permanente constituiu elemento chave para a obtenção dos resultados alcançados com a presente tese. À Prof.^a Georgete Medleg Rodrigues, coordenadora do Programa de Pós-Graduação em Ciência da Informação durante parte do período de desenvolvimento desta tese, quero expressar minha gratidão pelo apoio que recebi no programa de doutorado no PGCINF. Ao professor Sidney Barbosa, meu colega no Instituto de Letras da UnB, agradeço pelo “ombro amigo” nos momentos iniciais deste programa de doutoramento.

Je voudrais également remercier les professeurs de l'Université de Lille 3 (Charles de Gaulle) pour l'important appui intellectuel lors de mon séjour dans cette Université française. En particulier, j'ai un mot spécial de gratitude pour Stéphane Chaudiron, Directeur du Groupe d'Études et Recherche Interdisciplinaire en Information et Communication (GERiiCO), mon tuteur lors de mon séjour d'études à Lille (Doutorado Sanduíche). Je souhaite également remercier Joseph Mariani pour m'avoir mis en contact avec plusieurs groupes de recherche en France, ayant facilité mon séjour d'études à l'Université Lille 3. Je rends hommage aussi à Adama Samassékou, Président du Réseau Mondial pour la Diversité Linguistique (Réseau MAAYA), pour son dévouement à la cause du multilinguisme numérique et pour m'avoir permis de participer à plusieurs colloques internationaux sur ce thème. A mes anciens collègues à la Division pour la Société du Savoir de l'UNESCO, j'adresse mon chaleureux merci pour m'avoir introduit dans le monde du multilinguisme dans le cyberspace.

No tocante ao trabalho de programação do aplicativo em Python associado à metodologia aqui exposta, muito agradeço a Leandro Rodrigues. À equipe do LET On-Line (Socorro Lima, Cassio Nunes e Pedro Augusto da Silveira Bandeira) que facilitou e apoiou o processamento dos dados em computador, aqui registro a minha gratidão.

Quero ainda deixar inscrito o meu agradecimento aos meus familiares, que sempre me apoiaram e facilitaram meu trabalho durante o desenvolvimento do doutorado. À minha falecida mãe, a prof.^a Maria Vilani Sampaio de Menezes, pelo seu legado e exemplo de vida, sempre dedicada à educação brasileira, tendo guiado meus passos iniciais, despertando meu interesse pelo estudo e pela pesquisa.

Por fim, quero agradecer ao Estado Brasileiro, que financiou tanto meus estudos como os salários dos professores que me permitiram estudar, desde a escola primária até a obtenção do título de doutor em Ciência da Informação. A todos, muito obrigado.

RESUMO

A sociedade do conhecimento se caracteriza pela veloz ampliação do acesso a conteúdos digitais e pelo uso das tecnologias de informação e comunicação, as quais disponibilizam instantaneamente tais acervos em qualquer parte do globo. Mercê dessa nova ecologia tecnológica, significativas transformações ocorrem nos serviços oferecidos pelas bibliotecas e no seu relacionamento com usuários, tanto presencialmente como à distância. No contexto dessa atuação, além das fronteiras clássicas das bibliotecas e de uma necessidade de maior vitalidade linguística do português no mundo digital, a pesquisa objeto desta tese identifica alguns obstáculos ao acesso e compreensão de conteúdos científicos digitais por não falantes do português e propõe uma metodologia automatizada para facilitar a compreensão de textos científicos. A identificação das necessidades de falantes estrangeiros para acessar e compreender os conteúdos de bibliotecas digitais em língua portuguesa se fundamentou em dados coletados junto a estudantes estrangeiros francófonos, da Universidade de Brasília e da Universidade Charles de Gaulle (Lille 3) na França. O arcabouço conceitual interdisciplinar da tese está baseado em diversas disciplinas: arquitetura da informação, teoria da relevância, multimodalidade e no processamento da linguagem natural. A metodologia proposta no trabalho é operacionalizada por um aplicativo, desenvolvido em *software* livre na linguagem de programação Python, que integra a filtragem de texto, a sumarização textual automática, a tradução por máquina e o alinhamento sentencial automático. Esse dispositivo computacional pode ser utilizado em qualquer biblioteca interessada em ofertar um serviço personalizado para seus usuários estrangeiros. Os resultados obtidos são encorajadores, possibilitam uma maior vitalidade da língua portuguesa no campo científico e oferecem um campo fértil para novas pesquisas na áreas abordadas pela tese.

Palavras-chave: Acessibilidade. Alinhamento Sentencial. Biblioteca Digital. Multilinguismo. Multimodalidade. Serviços de Biblioteca. Sumarização Automática. Tradução por Máquina. Vitalidade Linguística.

ABSTRACT

The knowledge society is marked by the quick amplification of access to digital contents and by the use of information and communication technologies, which make available such contents everywhere in the globe. Thanks to this new technological ecology, significant transformations occur in services offered by libraries and in its relationship with users, both in face-to-face contact and at distance. In the context of such an activity beyond of the classical library frontiers and of a necessity of greater linguistic vitality of Portuguese in the digital world, this research identifies some obstacles for access and understanding of digital scientific contents by non-portuguese speakers and proposes an automated methodology to facilitate the understanding of scientific texts. The identification of foreign speakers' needs in accessing and understanding contents in digital libraries in Portuguese language is based on data collected with foreign francophone students, in Universidade de Brasília (in Brazil) and Université Chales de Gaulle (Lille 3, in France). The interdisciplinary conceptual framework of the thesis has its theoretical fundaments on several disciplines: information architecture, relevance theory, multimodality and natural language processing. The methodology proposed in this work is made operational through a set of computer programs, developed in the programming language Python under the concept of free software. It integrates four software devices: text filter, automatic text summarization, machine translation and automatic sentence alignment. This software can be used in any library that might be interested in offering a personalized service for its foreign users. The results achieved are very encouraging, ensure greater vitality of Portuguese language in the scientific field and offer a very fertile field for further research in this thesis subject.

Keywords: Accessibility. Automatic Sentence Alignment. Automatic Summarization. Digital Library. Library Service Personalization. Library Services. Linguistic Vitality. Machine Translation. Multilingualism. Multimodality. Scientific Information.

RESUMÉ

La société de la connaissance est marquée par un très rapide élargissement de l'accès à des contenus numériques et l'usage des technologies de l'information et communication, lesquelles mettent à disposition en ligne ce patrimoine numérique à toutes les parties du globe. Grâce à cette nouvelle écologie technologique, des transformations significatives ont lieu dans les services offerts par les bibliothèques et dans son rapport avec leurs usagers, soit à distance comme de façon présenteielle. Dans ce contexte, audelà des frontières classiques des bibliothèques et d'une nécessité d'une plus grande vitalité linguistique du portugais dans le monde numérique, la recherche objet de la présente thèse identifie quelques obstacles à l'accès et compréhension de contenus numériques par des non-parlants du portugais et on propose une méthodologie automatisée pour rendre plus facile la compréhension de textes scientifiques. L'identification des besoins des parlants étrangers pour accéder et comprendre les contenus de bibliothèques numériques en langue portugaise est basé sur des données collectés auprès des étudiants étrangers francophones, à l'Université de Brasilia e à l'Université Charles de Gaulle (Lille 3), en France. Le environnement conceptuel interdisciplinaire de cette thèse s'appuie en plusieurs disciplines: architecture de l'information, theorie de la pertinence, multimodalité et dans le traitement automatique de la langue. La méthodologie proposé dans ce travail est mis en fonctionnement par un outil informatique, développé en logiciel libre en langage Python, lequel intègre un filtrage du texte original, la sommairisation textuelle automatique, la traduction par machine e l'alignemenet sentenciel automatique. Ce dispositif computationnel peut être utilisé en n'importe quelle bibliothèque interessée pour offrir un service "à la carte" à ses utilisateurs étrangers. Les résultats obtenus sont très encourageants, permettent une plus grande vitalité linguistique de la langue portugaise dans le champ scientifique et offrent un terrain fertile pour des nouvelles recherches dans les domaines abordés para cette thèse.

Mots clés: **Accessibilité.** Alignement Sentenciel. Bibliotheque Numerique. Information Scientifique. Multilinguisme. Multimodalité. Services Bibliotecaires. Services de Bibliotheque à la Carte. Sommairisation Automatique. Traduction Automatique. Vitalité Linguistique.

“TOUT VIENT À POINT À QUI SAIT ATTENDRE”
(PROVÉRBIOS FRANCÊS)

LISTA DE FIGURAS

Figura 1 - Arquitetura da Informação.....	27
Figura 2 - Diagrama de um sistema geral de comunicação	31
Figura 3 - Modelo de Aceitação da Tecnologia de Thong, Hong e Tam.	58
Figura 4 - Diagrama conceitual de uma pesquisa científica	60
Figura 5 - Modelo Conceitual da Pesquisa.....	61
Figura 6 - Fluxograma do Dispositivo	85
Figura 7 - Página de Bi-Texto com Alinhamento Sentencial	86
Figura 8 - Exemplo de Bi-Texto com Identificação da Biblioteca Central da UnB	Erro!
Indicador não definido.	

LISTA DE QUADROS

Quadro 1 - Arquiteturas.....	28
Quadro 2 - Relevância de uma entrada para um indivíduo (*)	33
Quadro 3 - Sub-tarefas no processo global de compreensão Erro! Indicador não definido.	
Quadro 4 - Lista de Stop Sessions	77
Quadro 5 - Definição de Similaridade	83

LISTA DE TABELAS

Tabela 1 - Distribuição dos principais idiomas das páginas Web no domínio .gov.br	54
Tabela 2 - Trabalhos nos SNBUs de 2000-2008 sobre informações nas bibliotecas universitárias brasileiras geradas pelas tecnologias: categorias de análise	55

LISTA DE SIGLAS E ABREVIATURAS

AI	Arquitetura da Informação
ASA	Alinhamento Sentencial Automático
BD	Bibliotecas Digitais
BDTD	Biblioteca Digital de Teses e Dissertações
ClassSumm	Classification System
CPLP	Comunidade dos Países de Língua Portuguesa
FCF	Filtragem e Compatibilização de Formatos
FI	Filtragem da Informação
GistSumm	Gist Summarizer
IILP	Instituto Internacional da Língua Portuguesa
LC	Linguística Computacional
LSA	Latent Semantic Analysis
MDA	Multimodal Discourse Analysis
NEPPE	Núcleo de Ensino e Pesquisa de Português para Estrangeiros
NILC	Núcleo Interinstitucional de Linguística Computacional
NMT	Neural Machine Translation
PEC	Programa Estudantes Convênio
PLN	Processamento da Língua Natural
POSTRAD	Programa de Estudos de Tradução
RCAAP	Repositório Científico de Acesso Aberto de Portugal
ReGra	Revisor Gramatical
RI	Recuperação da Informação
SA	Sumarização Automática
SciELO	Scientific Electronic Library Online
SFG	Gramática Sistemico-Funcional
SFG	System Functional Grammar
SMT	Statistical Machine Translation
SR	Serviços de Referência
SuPor	Text Summarization in Portuguese
TA	Tradução Automática
TAC	Tradução Auxiliada por Computador
TAM	Modelo de Adoção de Tecnologias
TF-IDF	Term Frequency-Inverse Document Frequency
TF-IDF-Summ	Term Frequency-Inverse Sentence Frequency-based Summarizer

TM	Tradução por Máquina
TR	Teoria da Relevância
UnB	Universidade de Brasília
Unicamp	Universidade de Campinas
UNL	Universal Networking Language
W3C	World Wide Web Consortium
WDL	World Digital Library

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Multilinguismo no mundo digital e vitalidade linguística	19
1.2	Objetivo Geral	21
1.3	Objetivos Específicos	21
2	MARCO TEÓRICO	23
2.1	Arquitetura da Informação e Construção de Sentido	26
2.2	Contexto	27
2.3	Relevância em Ciência da Informação	29
2.4	Teoria da Relevância	30
2.4.1	Principais Pressupostos da TR	32
2.4.1.1	Relevância e Cognição	32
2.4.1.2	Relevância e comunicação	35
2.4.1.2.1	Princípio Comunicativo da Relevância	36
2.4.1.3	Relevância e compreensão	38
2.4.1.4	Arquitetura Mental da TR	40
2.5	Teoria da Relevância e Tradução	41
2.6	Multimodalidade	42
2.6.1	A CI, a Arquitetura da Informação e a Multimodalidade	43
2.7	Tecnologias das Línguas	47
2.8	Considerações sobre a internacionalização de Bibliotecas Digitais	54
2.8.1	Contextualização	54
2.8.2	A Biblioteca Híbrida e os Serviços de Referência	56
2.8.3	Usabilidade e Internacionalização de Bibliotecas Digitais	57
3	METODOLOGIA DA PESQUISA	59
3.1	Elementos Conceituais	59
3.2	Descrição da metodologia da pesquisa	60
3.3	Implementação da Metodologia	63
3.4	Coleta e Análise dos Dados	64
3.4.1	Coleta de Dados	64
3.4.2	Coleta e Análise dos Dados	66
3.4.2.1	Questionário 1 – Identificação	66
3.4.2.2	Questionário 1 – Questões Gerais	66
3.4.2.3	Questionário 1 – Questões sobre o site web da Universidade de Brasília	67

3.4.2.4	Questionário 1 – Questões Técnicas.....	68
3.4.2.5	Questionário 2 – Identificação	69
3.4.2.6	Questionário 2 – Questões Gerais	69
3.4.2.7	Questionário 2 – Questões sobre recursos digitais de informação científica em Português.....	70
3.4.2.8	Questionário 2 – Questões Técnicas.....	71
4	ACESSO E COMPREENSÃO DE CONTEÚDOS EM PORTUGUÊS POR ESTRANGEIROS EM BIBLIOTECAS DIGITAIS CIENTÍFICAS: METODOLOGIA PROPOSTA.....	73
4.1	Contexto.....	73
4.2	Usabilidade e Internacionalização de Websites	73
4.3	Componentes da Proposta da Tese	75
4.4	Aplicativo Computacional para Facilitar a Compreensão do Texto da Biblioteca Digital Científica	76
4.5	Disseminação da Metodologia por Bibliotecas	Erro! Indicador não definido.
4.6	Uso Remoto da Metodologia.....	90
5	BIBLIOTECAS DIGITAIS E REPOSITÓRIOS PARA APLICAÇÃO PRÁTICA DA METODOLOGIA	91
6	CONSIDERAÇÕES FINAIS	93
	REFERÊNCIAS BIBLIOGRÁFICAS	96
	ANEXO A – PRÉ-TESTE – QUESTIONNAIRE 1	104
	ANEXO B – PRÉ-TESTE.....	109
	ANEXO C – PRÉ-TESTE – TABULAÇÃO DO QUESTIONÁRIO 1 (TABULATION DU QUESTIONNAIRE 1).....	114
	ANEXO D – QUESTIONÁRIO 2.....	130
	ANEXO E – TABULAÇÃO E GRÁFICO DO QUESTIONÁRIO 2.....	138
	ANEXO F – TABULAÇÃO DO QUESTIONÁRIO 2 – VERSÃO 2ERRO! INDICADOR NÃO DEFINIDO.	
	ANEXO G – TUTORIAL PARA USO DO APLICATIVO COMPUTACIONAL	181
	ANEXO H – BI TEXTO DE DISSERTAÇÃO DE MESTRADO	183
	ANEXO I – BI-TEXTO DE UMA TESE DE DOUTORADO	187
	ANEXO J – LICENÇAS DE USO DE SOFTWARE.....	197

1 INTRODUÇÃO

O surgimento da sociedade de rede, no dizer de Castells (2008) e de autores como Daniel Bell (1976), Marc Porat (1977), Michael Hardt e Antonio Negri (2000), Peter Drucker (2001), Hardt *et al.* (2002) e Franck Webster (2006), deu origem a diversos novos fenômenos em todas as dimensões da atividade humana. Um desses novos fenômenos se denomina multilinguismo no mundo digital, tema no qual o acesso à informação científica se inscreve e se desenvolve a presente tese de doutorado.

1.1 Multilinguismo no mundo digital e vitalidade linguística

A partir da convergência das tecnologias de tratamento de textos, imagens e sons para o paradigma digital e com a criação dos protocolos de comunicação e do hipertexto, diversas transformações ocorreram no tocante ao acesso à informação. Está, portanto, em curso “uma reestruturação de ideias, comportamentos e conceitos na sociedade, gerando uma nova dinâmica em decorrência, principalmente, da crescente necessidade de utilização de informações” (LAZZARIN *et al.*, 2012, p. 232).

A primeira dessas ideias corresponde a uma crescente eliminação de fronteiras para acesso a conteúdos digitais (desterritorialização), seja sob a forma de texto, de som ou de imagem. A supressão das fronteiras de acesso a conteúdos digitais facultada pela Internet e pelo hipertexto garante novos espaços informacionais ao usuário situado à distância do local em que se encontram fisicamente tais conteúdos. Esse processo dá origem a novos fenômenos resultantes da ausência do contato presencial do usuário com as pessoas que trabalham nas instituições mediadoras de informação. Por outro lado, abre uma ampla perspectiva para novas modalidades de interação a distância, assim como para novos serviços ofertados no mundo digital por tais instituições.

De particular interesse tivemos a mudança de paradigma relativo ao contato entre falantes de línguas diferentes. Presenciamos cada vez mais uma intensa atividade de línguas em contato na web, fenômeno que anteriormente só acontecia nas regiões fronteiriças de países com falantes de idiomas diferentes.

Nesse contexto, tornou-se comum a disponibilização de conteúdo e do mesmo texto em diversas línguas, através da Internet, principalmente por meio de *websites* multilíngues. Um exemplo bem ilustrativo desse novo evento tecnológico foi a criação da *World Digital Library - WDL* (<http://www.wdl.org>), iniciativa da UNESCO para disponibilizar no maior número de línguas possível, conteúdos de interesse da humanidade nos campos da educação, ciência, cultura e comunicação. Atualmente, as línguas dos conteúdos da WDL são: árabe, chinês, espanhol, francês, inglês, português e russo. Outra iniciativa relevante, a biblioteca digital *Europeana* (<http://www.europeana.eu>), oferece acesso a mais de 54 milhões de obras de arte, objetos, livros, vídeos e artefatos sonoros de toda a Europa, com possibilidade de acesso em 24 línguas da União Europeia.¹

Não menos importante, a expansão das capacidades de armazenamento de dados e informações sob a forma digital propicia oferta de gigantescos volumes de dados em suportes diferentes, formatos variados e organização institucional inovadora. A disponibilidade de conteúdos de múltiplas naturezas e gêneros também enseja a possibilidade de consulta por falantes de línguas diferentes daquela do conteúdo original acessível pelas redes de informação em linha.

Vale realçar ainda o avanço do processamento da língua natural (PLN) e de outras tecnologias da língua, cujo desenvolvimento, embora um pouco tardio e mais demorado, se comparado com o tratamento computacional em outros campos do conhecimento, começou a se tornar presente na sociedade do conhecimento a partir da década de 1950.

O fenômeno da vitalidade linguística também repercute no mundo digital de forma bastante expressiva. Se antes do processo de mundialização já havia uma marcante preocupação geopolítica de muitos países na difusão das suas línguas, o fenômeno se acentua no mundo digital. Seja no campo empresarial, seja no campo científico, tornou-se frequente a criação de *websites* multilíngues ou ainda de *interfaces* informativas multilíngues, de forma a facilitar o acesso aos conteúdos por audiências cada vez mais amplas e de origem linguística diversificada. No campo educacional e científico, numerosas universidades e bibliotecas atendem usuários de diversas línguas e por isso seus *websites* são projetados com interfaces

¹ Acessos à WDL e à Europeana em 29.mar.2017

multilíngues. Na esfera da literatura científica em português, a Biblioteca Digital de Teses e Dissertações (BDTD) (<http://bdttd.ibict.br>), uma iniciativa brasileira, conta com a participação de 105 instituições e disponibiliza 476.764 documentos (128.125 teses e 345.639 dissertações) e oferece *interfaces* em português do Brasil, inglês e espanhol. Iniciativas como OASIS (Portal Brasileiro de Publicações Científicas em Acesso Aberto) (<http://oasisbr.ibict.br>) e o *Scientific Eletronic Library Online* (SciELO) (<http://www.scielo.org>) também merecem destaque. Em Portugal, por sua vez, o Repositório Científico de Acesso Aberto de Portugal (RCAAP) (<https://www.rcaap.pt/>) garante acesso a 1.321.231 documentos.²

A reunião de todos esses elementos no novo ambiente tecnológico motivou a presente pesquisa, objetivando a formulação conceitual e implementação de uma metodologia de acesso por estrangeiros a conteúdos científicos de bibliotecas digitais em português.

O tema da vitalidade linguística se insere nas práticas de políticas linguísticas e caracteriza uma preocupação cada vez maior da comunidade de falantes de cada língua, das suas academias de língua e de outras instituições linguísticas. Por ser de interesse mútuo, imaginamos pertinente considerarmos a análise da situação de pares de línguas no mundo digital, a fim de permitir uma maior cooperação entre comunidades linguísticas diferentes, com benefício recíproco da vitalidade linguística de cada uma delas. Daí, a nossa formulação do objetivo geral e dos objetivos específicos desta tese, como se apresenta a seguir.

1.2 Objetivo Geral

Propor uma metodologia destinada a facilitar o acesso e a compreensão de conteúdos científicos em bibliotecas digitais em português por estrangeiros.

1.3 Objetivos Específicos

Apresentar recomendações técnicas destinadas a facilitar o acesso à literatura científica em língua portuguesa por estrangeiros disponibilizada em repositórios de bibliotecas digitais.

² Acessos à BDTD e RCAAP em 28.03.2017

Propor o uso de tecnologias linguísticas em bibliotecas digitais científicas para facilitar a compreensão rápida de conteúdos em português por estrangeiros;

Integrar os processos estudados em uma metodologia prática com ferramentas computacionais desenvolvidas para seu uso, a ser disseminada a partir de ambientes físicos e virtuais de referência e disponível para uso local ou remoto por parte de estrangeiros.

2 MARCO TEÓRICO

O acesso à informação geralmente é intermediado por técnicas que tomam em consideração o perfil do usuário, visto que os sistemas e mecanismos de intermediação do acesso à informação são projetados para melhor atender a necessidades específicas de cada perfil. No entanto, é de se notar que no mundo digital se torna bem mais complexa a gestão de perfis de usuário, na medida em que o acesso ao conteúdo informacional pode ser feito a distância, sem que haja contato pessoal com os especialistas da unidade de informação. Sobre a criação de perfis de usuário no mundo digital, alguns estudos ainda preliminares abordam o assunto (GOLEMATTI *et al*, 2007; TRAJKOVA; GAUCH, 2004).

A personalização do serviço oferecido por bibliotecas digitais começa a se tornar uma preocupação crescente no campo acadêmico e na prática da biblioteconomia, por se tratar de fenômeno relativamente novo e que requer estudos mais acurados. No caso de usuários estrangeiros, a crescente internacionalização das universidades reclama um tratamento específico para essa categoria de usuário, ainda pouco estudada na literatura científica. Essa categoria está cada vez mais presente na vida das bibliotecas, visto que é comum e cada vez mais crescente a presença de estudantes estrangeiros nas universidades. Além disso, a disponibilidade trans-fronteiras dos conteúdos das bibliotecas digitais impõe a necessidade de criação de uma personalização de serviços das unidades de informação para tais usuários.

No entanto, essa preocupação ainda não deu lugar a pesquisas de maior profundidade sobre o perfil de um usuário estrangeiro que busca informação em uma biblioteca digital. Com efeito, os trabalhos acadêmicos têm se voltado principalmente para estudos do usuário e da necessidade de informação. Em artigo intitulado *On user studies and information needs* (WILSON, 1981), no qual conceitos tais como “universo do conhecimento” e um modelo de comportamento informacional são desenvolvidos, não há análises sobre o relacionamento de um usuário remoto com bibliotecas digitais e seus serviços. No mesmo contexto, o comportamento de um usuário estrangeiro de uma biblioteca digital também não faz parte do “universo do conhecimento”, nem muito menos do modelo de busca e necessidade de informação apresentado por tais autores.

Igualmente frequentes são estudos sobre o perfil do usuário objetivando oferecer-lhe sistemas de informação de acesso personalizado. Nesse tipo de estudo, em geral é feita uma coleta de dados sistemática a partir de arquivos (*logfiles*), com o objetivo de conhecer o perfil do usuário, a partir do qual são oferecidos serviços de informação personalizados. Com tal propósito, o artigo *User Profiles for Personalized Information Access* (GAUCH *et al*, 2007) discute de forma detalhada como construir perfis de usuários a partir de diversas técnicas (frequência de palavras, redes semânticas e outros critérios de ranqueamento), com o objetivo de propor o desenvolvimento de sistemas de acesso personalizado.

Uma outra abordagem para estudo do binômio biblioteca digital versus usuários pode ser encontrada em pesquisas para modelagem automática do usuário para bibliotecas digitais personalizadas. Trata-se aqui de “descobrir” o perfil do usuário por meio de uma análise de dados obtida ora automaticamente, ora por meio de um cadastro de dados de cada usuário. Esse enfoque também analisa uma arquitetura genérica de uma biblioteca digital adaptável aos seus distintos usuários. O artigo *Automated User Modelling for Personalized Digital Libraries* (FRIAS-MARTINEZ *et al*, 2006) oferece uma descrição bastante acurada tanto do perfil do usuário como dos possíveis serviços que a biblioteca digital poderá oferecer em busca de uma personalização para diferentes clientelas. Esse artigo indica que uma biblioteca digital se constitui via de regra de quatro componentes: 1) Informação; 2) Estrutura (descrevendo as características sintática e a semântica da informação oferecida pela BD; 3) Elementos de interação (onde é abordada a interface de pesquisa, o desenho das telas); e 4) as propriedades (onde são tratadas as questões de segurança, direito de autor sobre as informações oferecidas pela BD). Os serviços oferecidos pela BD através dos elementos de interação apresentados no artigo se classificam em três grupos, a saber:

Mecanismos para a personalização de conteúdo: tornam possível para cada usuário criar uma BD pessoal que contém somente a informação de interesse para si;

Mecanismos para auxiliar no processo de navegação: oferecem a cada usuário um ambiente que melhor se adequa à maneira pela qual a interação com a BD é exercida;

Mecanismos de Filtragem da Informação (FI) e Recuperação da Informação (RI): serviços que proporcionam maneiras de encontrar e filtrar a vasta quantidade de informação que o usuário acessa e recebe.

No entanto, quando se trata de examinar o relacionamento de um usuário estrangeiro, remoto ou local, com o acervo e os serviços oferecidos por bibliotecas digitais, a literatura escasseia, para não dizer que quase inexistente, visto que o fenômeno não tem sido ainda muito abordado com propriedade em trabalhos acadêmicos.

Feitas tais ponderações, consideramos ser importante identificar elementos conceituais relativos aos serviços que podem ser oferecidos ao usuário estrangeiro para facilitar seu acesso e compreensão dos conteúdos científicos de tais bibliotecas e constituir uma metodologia com essa finalidade. Vale, no entanto, observar que – embora este não seja o foco principal da tese – há necessidade de fazer algumas considerações sobre o projeto de *website* e seu *design*, o que nos remete principalmente a questões relacionadas à arquitetura da informação. Quanto aos serviços a serem oferecidos, o foco principal naturalmente se vincula ao tratamento automático da língua e em particular a ferramentas de auxílio à tradução³. É portanto com base nesses elementos que buscamos identificar marcos teóricos que dão suporte à proposta metodológica da tese.

Nessa perspectiva, pressupõe-se que tais elementos conceituais oferecem algumas respostas para a questão central da tese: a busca e compreensão de conteúdos científicos de bibliotecas digitais em português por estrangeiros.

³ Ferramenta de tradução: “ferramenta de informática que contribui para a realização de uma atividade ligada à <tradução>.”

Nota 1 – Os programas de <tradução assistida por computador> mais comuns são: os bi-textos (aqueles que apresentam os <textos de partida> e os <textos de chegada> em duas colunas), também conhecidos como <corpora paralelos> ou <textos paralelos>; os <programas de extração automática de termos> ou <concordanciadores>, que permitem o armazenamento de textos e buscam palavras ou expressões em um corpus monolíngue ou bilíngue; os dicionários eletrônicos; os bancos de dados terminológicos; as <memórias de tradução> (programas de armazenamento do <texto de partida> e de <texto de chegada> em duas bases de dados paralelas, assim como de sua sugestão de tradução cujo haja semente semelhante no banco de dados constituído).

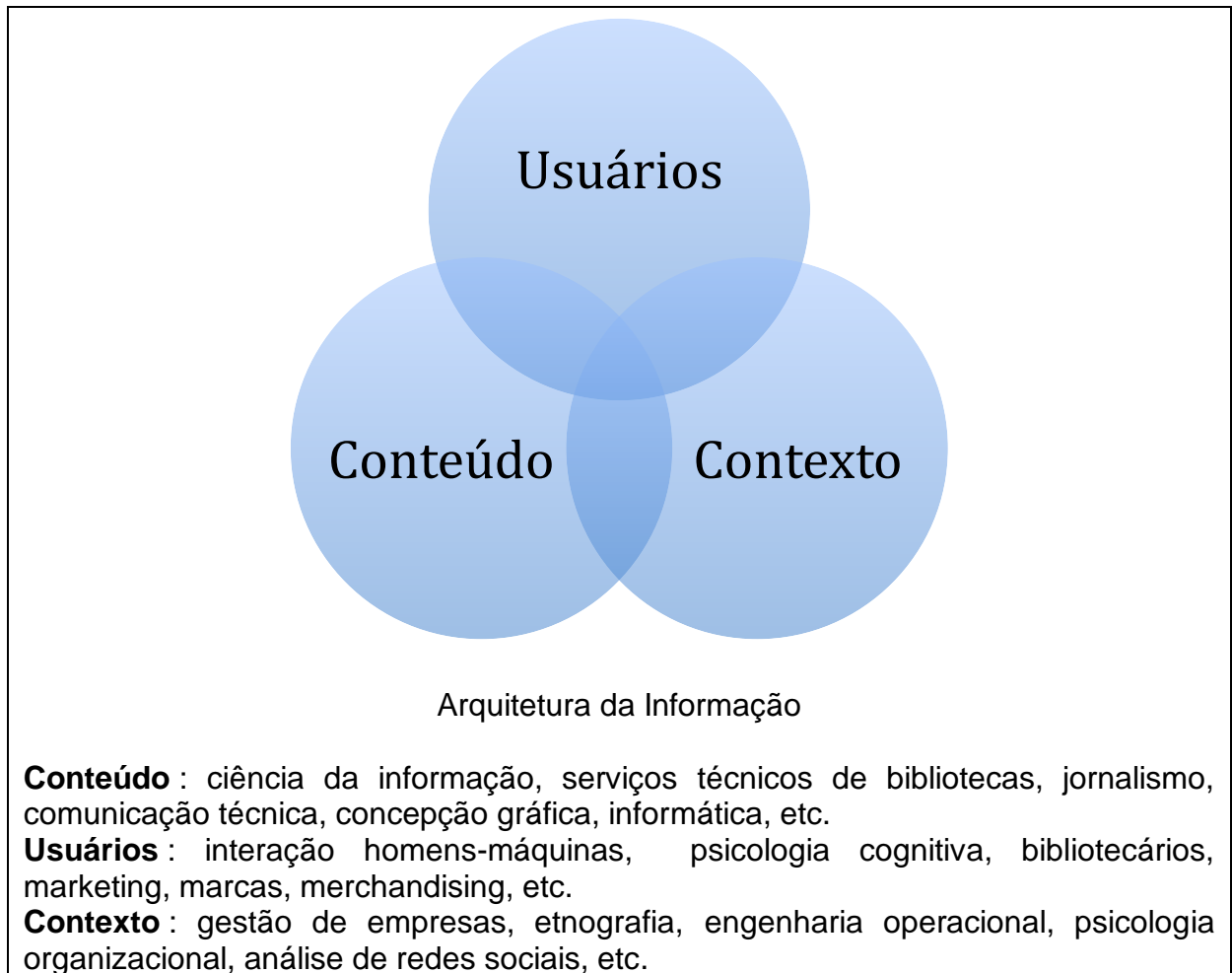
Nota 2 – Além das ferramentas de tradução, o tradutor dispõe de outros programas que apresentam funções de caráter linguístico como a contagem de palavras, os corretores ortográficos e gramaticais, o revisor de texto.” (LEE-JAHNKE *et al.*, 2013).”

2.1 Arquitetura da Informação e Construção de Sentido

O pioneirismo no uso do termo “Arquitetura da Informação” é atribuído ao arquiteto Richard Saul Wurman, que assim definiu “Arquiteto da Informação”: ” (1) o indivíduo que organiza os padrões inerentes de dados, tornando claro o que seria complexo; (2) uma pessoa que cria a estrutura ou mapa de informação que permite aos outros encontrar seus caminhos pessoais para o conhecimento (grifo nosso); (3) a ocupação profissional emergente no século XXI abordando as necessidades da época, focado na clareza, entendimento humano e na ciência da organização da informação” (WURMAN *apud* COOPER, 2009, p. 2).

Outros estudiosos introduziram conceitos inerentes à arquitetura da informação, como é o caso de Andrew Hinton (2009), para quem o conceito de “contexto” é primordial quando se trata de arquitetura de informação. Uma tentativa de reunir esses conceitos nos é oferecida por Broudoux, Chartron e Chaudiron (2013), como ilustrado na figura 1.

Figura 1 - Arquitetura da Informação



Fonte: Broudoux, Chartron e Chaudiron (2013, p. 16)

2.2 Contexto

Em seu artigo *The Machineries of Context* (HINTON, 2009, p. 42), Andrew Hinton pontua que “a web criou um complexo amálgama de documentos e edifícios. Com documentos normais em papel, nós os lemos, os arquivamos, os jogamos fora ou os enviamos para alguém. Nós não ‘vamos aos documentos’. Nós não os visitamos. Documentos web são diferentes. Eles são **lugares** na web”. Para Hinton, arquitetura da informação é a arquitetura para esse “complexo amálgama”, um tipo de espaço, ou seja, um contorno de experiência contextual obtida pela criação de fronteiras e conexões. Uma nova espécie de arquitetura tornou-se necessária, visto que para Hinton a *web* é e tem sido sempre social e por isso são necessários mecanismos e estruturas – a maquinaria do contexto – para assegurar que uma atividade possa ocorrer na web. E por isso são requeridas regras e paradigmas.

Poderíamos pensar, portanto, em uma primeira aproximação da arquitetura da informação (AI) para o nosso problema de estudo: uma AI para *websites* em português, cujo contexto deveria ser também adaptado ao falante de língua materna estrangeira. O Quadro nº 1 ilustra bem o conceito de contexto, no sentido formulado por Hinton.

Quadro 1 - Arquiteturas

(Jovens Egípcios) olhando atentamente através das janelas da Internet, ganharam um mais agudo sentido do que muitos dos seus antepassados sobre as liberdades e oportunidades que aqueles não tiveram. Encontraram na mídia social um meio de interagir e compartilhar ideias, contornando, no espaço virtual, as restrições colocadas pela liberdade de reunião física. (Mohamed ElBaradei, ex-Diretor da Agência Internacional de Energia Atômica, das Nações Unidas, *New York Times*, 10 de Fevereiro de 2011).

Fonte: Arango, J. (2011, p. 41)

Da mesma forma que há uma insuficiência de pesquisas para definir perfis de usuário na ecologia tecnológica da sociedade do conhecimento, fenômeno análogo ocorre no tocante a uma formulação teórica para dar embasamento à questão do acesso por usuários estrangeiros a sites em outras línguas. Dentre as diversas formulações teóricas cuja viabilidade poderia ser considerada para tratar esse tema, foram selecionados alguns conceitos que poderão permitir:

- a) estabelecer uma formulação para evitar que o usuário desista da sua busca;
- b) obtida a “fidelidade” mencionada no item anterior, analisar o fenômeno da compreensão do conteúdo, por meio de ferramentas que facilitem a compreensão desses conteúdos.

Com relação ao processo de obtenção da “fidelidade” do usuário a partir do acesso a um website, apresenta-se a seguir uma discussão sobre a Teoria da Relevância, que poderá contribuir para a adaptação de *websites* de bibliotecas digitais às necessidades de usuários estrangeiros

A literatura sobre relevância está contemplada por duas principais correntes de estudo: 1) a abordagem do conceito de relevância sob a ótica do usuário e da

ciência da informação; e 2) a concepção cognitiva e comunicacional da Teoria da Relevância (popularizada principalmente pelos estudos de Deirdre Wilson e Dan Sperber (2002)); neste caso, o conceito de relevância está associado à pragmática, parte da linguística descrita como o estudo do uso da língua (em oposição ao estudo da sua estrutura).

2.3 Relevância em Ciência da Informação

Sem estar inicialmente ligado à teoria da relevância (TR), o conceito de relevância tem sido ao longo do tempo estudado em ciência da informação por diversos autores (HILLMAN, 1964; SARACEVIC, 1975, 2007; entre outros). Há no entanto, uma discussão ainda não conclusiva sobre uma **definição** desse conceito na ciência da informação sobre o qual os cientistas da informação ainda não lograram chegar a um consenso.

No artigo *Relevance: The Search for a Definition*, Linda Schamber e Michael Eisenberg, da Syracuse University (1988), fazem uma revisão crítica do tema sob a ótica da ciência da informação no tocante às abordagens anteriormente apresentadas: relevância orientada por sistema, orientada por usuário, multidimensional e cognitiva. A partir dessa revisão crítica, Schamber e Eisenberg introduzem o conceito de relevância centrada no usuário, baseado em três postulados da relevância em CI, a saber: a) é conceito multidimensional baseado em um processo de julgamento humano; b) depende tanto de fatores internos (cognitivos) e externos (situacionais); e c) é um conceito intersubjetivo mas ainda assim sistemático e mensurável. Embora esse enfoque não tenha uma ligação mais direta com o que veio a constituir a teoria da relevância, a abordagem centrada no usuário abandona o paradigma clássico do modelo de comunicação, baseado exclusivamente no binômio fonte-destinatário. Trata-se de um modelo cognitivo que coloca o usuário – e não o sistema – como elemento central e determinante das dimensões de relevância, como veio a ocorrer posteriormente com a teoria da relevância apresentada brevemente a seguir.

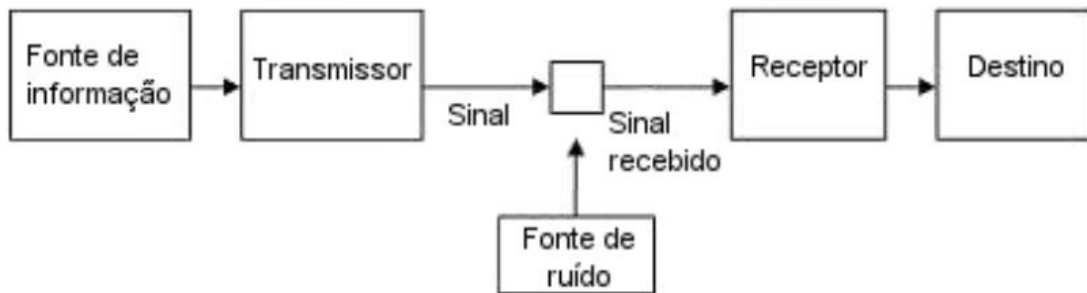
2.4 Teoria da Relevância

A escolha de signos a serem colocados na página de abertura de um site web é de fundamental importância para assegurar a permanência do usuário nesse site, assim evitando sua desistência de buscar a informação procurada, seja diretamente ou fazendo uso de serviços oferecidos pelas bibliotecas. Com maior razão, isto ocorrerá quando o usuário não compreende integralmente o idioma em que estiver escrito esse conteúdo. Trata-se portanto de um processo comunicacional entre a página *web* e seu usuário.

A afirmação central da Teoria da Relevância (TR) é a de que expectativas de relevância são precisas e previsíveis o suficiente para guiar o destinatário do texto na direção do significado proposto pelo comunicador. Como no escopo desta tese o destinatário da *webpage* em português de uma biblioteca digital será um estrangeiro, um *design* específico para esse tipo de usuário poderia eventualmente ser considerado. Com base em Sperber e Wilson (2002), o objetivo é explicar em termos cognitivamente realísticos a que equivalem essas expectativas e como elas podem contribuir para uma abordagem empiricamente plausível da compreensão. Para tal propósito, delineiam-se as principais suposições da versão atual da teoria e discutem-se algumas de suas implicações. Os conceitos da TR são apresentados por Sperber e Wilson nos seguintes tópicos: Relevância e Cognição, Relevância e Comunicação, Relevância e Compreensão e Relevância e Arquitetura Mental.

A TR se propõe a desenvolver em detalhe um dos principais argumentos de Grice, na área da comunicação humana. Para Grice, um elemento essencial da comunicação humana – tanto verbal como escrita – é a expressão clara e o reconhecimento de intenções (GRICE, 1989). Os conceitos desenvolvidos nesse campo questionaram o entendimento clássico proveniente da teoria da comunicação (SHANNON; WEAVER, 1949, 2015) pelo qual um comunicador codifica um sinal a ser decodificado ao ser percebido pela audiência, como mostra a figura 2.

Figura 2 - Diagrama de um sistema geral de comunicação



Fonte: Infopedia.pt, disponível no Google Images

No modelo inferencial de comunicação, o emissor estabelece evidência da sua intenção de conduzir um certo significado, que por sua vez pode ser **inferido** pela audiência com base na evidência que lhe foi oferecida pelo comunicador. A meta da pragmática inferencial consiste em compreender COMO um receptor (*hearer*) INFERE o significado que lhe foi submetido por um emissor (*speaker, communicator*). Esta abordagem teórica repousa também noutra assertiva de Grice, indicando que uma **sequência de palavras** (*utterance*) cria automaticamente expectativas muito favoráveis (*expectations*) que irão guiar o receptor em torno do significado submetido pelo emissor. Tais expectativas são descritas por Grice em termos de um Princípio Co-operativo e de “máximas” de qualidade (veracidade, *truthfulness*), quantidade (informatividade, *informativeness*), relação (relevância, *relevance*) e maneira (clareza, *manner*) que os emissores são instados a observar. Nesses termos, a interpretação que um *receptor racional* irá escolher será aquela que melhor satisfaz à sua expectativa favorável. A assertiva central da TR pressupõe que as **expectativas muito favoráveis** (*expectations*) levantadas por uma sequência de palavras sejam suficientemente precisas e previsíveis para guiar o receptor para compreender o significado que o emissor pretende apresentar.

No relacionamento entre o usuário (*receptor*) e o *website* (*emissor, comunicador*), a evidência oferecida pelo comunicador deverá estar presente na página de abertura do site e nas seguintes, permitindo que o usuário estrangeiro distante possa confiar que irá encontrar a informação por ele buscada ou recorrer a algum serviço oferecido pela biblioteca digital.

2.4.1 Principais Pressupostos da TR

Neste tópico serão explicitados a terminologia e conceitos básicos sobre os quais se apoia a TR e que serão oportunamente mencionados nas considerações sobre *websites*, tema complementar à metodologia que se desenvolveu nesta tese. Os conceitos da TR são agrupados por Sperber e Wilson em quatro categorias e são brevemente desenvolvidos nos seguintes tópicos: Relevância e Cognição, Relevância e Comunicação, Relevância e Compreensão e Relevância e Arquitetura Mental.

2.4.1.1 Relevância e Cognição

Wilson e Sperber (2002) consideram que relevância é uma propriedade potencial não só de uma sequência de palavras (*utterance*) e de outros fenômenos observáveis, mas de pensamentos, recordações e conclusões inferidas. Nos termos da TR, qualquer estímulo externo ou representação interna que alimenta um processo cognitivo pode ser relevante para um indivíduo em algum momento. De acordo com essa teoria, uma sequência de frases suscita expectativas favoráveis de relevância porque a busca pela relevância é um atributo básico da cognição humana, explorável pelos comunicadores (e não porque comunicadores possam obedecer a um Princípio de Co-operação, ou a máximas, ou a algum tipo de convenção específica de comunicação). Cabe portanto explicar a noção básica cognitiva de relevância e o Princípio Cognitivo da Relevância, fundadores da abordagem teórica da relevância, ou seja, a pragmática.

Primeiramente, indaga-se: quando uma entrada é relevante? Intuitivamente, uma entrada (uma imagem vista, um som, uma sequência de palavras, uma recordação, um insumo) é relevante para um indivíduo quando for conectada a uma informação anterior (disponível para o receptor produzir conclusões que façam sentido para si: por exemplo, oferecendo resposta a uma questão que se tenha em mente, melhorando conhecimento sobre um certo tópico, esclarecendo uma dúvida, confirmando uma suspeita ou corrigindo uma impressão equivocada). Na TR, uma entrada (ou um insumo para um sistema de recepção) é relevante para um receptor quando o seu processamento em um contexto de pressuposições aceitas produz um **EFEITO COGNITIVO POSITIVO**, entendido como uma diferença importante para a

representação que o indivíduo faz do mundo – uma verdadeira conclusão, por exemplo. Falsas conclusões, embora tenham um efeito cognitivo, não interessam por não oferecer efeitos cognitivos positivos.

Um outro conceito oferecido pela TR considerado importante é a **implicação contextual**, uma conclusão dedutível da entrada (insumo) e do contexto juntos, mas nunca isoladamente. De acordo com a TR, uma entrada é RELEVANTE para um indivíduo quando e somente quando seu processamento gera um EFEITO COGNITIVO POSITIVO.

Relevância também não pode ser tomada de forma absoluta, deve ser vista como um conceito gradual. A TR defende que um *input* torna-se meritório para ser considerado entre outros elementos não somente por sua relevância, mas pelo fato de ser mais relevante do que qualquer outro *input* disponível no momento. De forma intuitiva, quanto maior for o esforço de percepção, memorização e inferência requerido, menos apropriado ao processamento será esse *input* e, conseqüentemente, menos chamador da nossa atenção. Resumidamente, conforme o quadro 2, outros elementos sendo iguais, quando maior for o **esforço de processamento** requerido, menos relevante esse *input* será e portanto o conceito de **relevância** pode ser aprofundado em termos de efeitos cognitivos e esforço de processamento.

Quadro 2 - Relevância de uma entrada para um indivíduo (*)

- a) quanto maiores forem os **efeitos cognitivos positivos** obtidos pelo processamento dessa entrada, maior a sua relevância para um indivíduo naquele momento;
- b) quanto maior for o **esforço de processamento** dispendido, menor será a relevância dessa entrada para um indivíduo naquele momento.

(*)(mantidos iguais os outros elementos do contexto)

O exemplo conceitual a seguir ilustra os aspectos da relevância acima citados.

“Maria, que não aprecia carne e é alérgica a galinha, liga para um fornecedor de jantar para conhecer a oferta do menu. Ela obtém as seguintes respostas:

- (1) Estamos servindo carne;
- (2) Estamos servindo galinha;

- (3) Ou servimos galinha ou $(7^2 - 3)$ não é 46.” (WILSON & SPERBER, 2005, p. 225).

De acordo com a caracterização de relevância, as 3 sequências de palavras são relevantes para Maria, mas (2) é mais relevante do que (1) e (3). É mais relevante do que (1) por razões de efeito cognitivo: (2) engloba (1) e, em consequência, produz todas as consequências derivadas de (1). Será também mais relevante do que (3) por razões de esforço de processamento: embora (2) e (3) sejam logicamente equivalentes, e portanto produzam os mesmos efeitos cognitivos, tais efeitos são mais facilmente derivados de (2) do que de (3), que requer um esforço adicional de *parsing* de inferência (pela constatação de que a segunda disjuntiva é falsa e a primeira é verdadeira). Pode-se então concluir que (2) é a mais relevante sequência de palavras para Maria, por razões de esforço e efeito. De modo geral, pode-se afirmar que, quando uma quantidade similar de esforço é requerida, o fator **efeito** é decisivo para determinar os graus de relevância e quando similar esforço de efeito é necessário, o fator **esforço** é decisivo na determinação do **grau de relevância**.

A caracterização da relevância é predominantemente qualitativa e não quantitativa, como se observou no exemplo acima. Embora também possa ser útil explorar as dimensões quantitativas da relevância, de um ponto de vista formal, é a noção qualitativa em lugar da noção quantitativa que provavelmente irá oferecer um melhor ponto de partida para a construção de uma teoria psicologicamente plausível. A TR advoga com ênfase que os seres humanos têm uma tendência automática para maximizar a relevância, não porque tenhamos uma escolha na matéria – raramente a temos – mas pela forma pela qual nosso sistema cognitivo evoluiu. Resultante de uma constante pressão em torno de uma crescente eficiência, o sistema cognitivo humano se desenvolveu de tal forma que o nosso mecanismo de percepção tende automaticamente a selecionar estímulos potencialmente relevantes. Da mesma forma, nosso mecanismo de recuperação de memória tende automaticamente a ativar pressupostos potencialmente relevantes e nosso mecanismo inferencial tende espontaneamente a processá-los da forma mais produtiva. Em termos de uma formulação, a TR assim enuncia o Princípio Cognitivo da Relevância: A cognição humana tende a ser ajustada para a maximização da relevância (WILSON; SPERBER, 2005, p. 227).

2.4.1.2 Relevância e comunicação

De acordo com a TR, há uma tendência cognitiva universal para maximizar a relevância, tornando possível prever e manipular o estado mental de outras pessoas, até um certo nível. Conhecendo essa tendência humana de selecionar o estímulo mais relevante no nosso ambiente e processá-lo de forma a maximizar a sua relevância, pode-se estar apto a produzir um estímulo com capacidade de atrair a atenção do interlocutor, de habilitar a recuperação de certos pressupostos contextuais e de orientá-lo para uma conclusão pretendida. Há diversos exemplos na produção desse estímulo para, de certa forma, induzir um interlocutor ou uma audiência a adotar uma ação desejada pelo emissor. A TR a esse propósito apresenta a noção de **COMUNICAÇÃO INFERENCIAL-OSTENSIVA**, que pode ser decomposta em:

- a) Intenção informativa (a intenção de informar a uma audiência sobre algo);
- b) Intenção comunicativa (a intenção de informar a uma audiência de uma intenção informativa de alguém).

Para Wilson e Sperber (2004), o entendimento é alcançado quando a intenção comunicativa é preenchida – isto é, quando a audiência reconhece a intenção informativa (Se a intenção informativa por sua vez cumpre seu papel, dependerá do nível de confiança que a audiência tiver no comunicador. Há uma lacuna entre entendimento e crença. Para o entendimento ser alcançado, a intenção informativa deve ser reconhecida, mas não tem obrigatoriamente que ser efetivada).

As questões de intencionalidade no que se refere à comunicação inferencial-ostensiva envolvem o uso de um **ESTÍMULO OSTENSIVO**⁴, projetado para atrair a atenção de uma audiência, com foco no significado do comunicador. A TR advoga que o uso de um estímulo ostensivo pode criar expectativas fortes (*expectations*), precisas e previsíveis de relevância não obtidas por outros estímulos. A descrição dessas expectativas fortes e como elas podem auxiliar a audiência na identificação do significado que se quer comunicar também é desenvolvida na TR.

⁴ Uma das autoras consultadas, Longchamps (2014), prefere usar o termo “ostensão” em lugar de “estímulo ostensivo”. Nesta tese, mantivemos o termo “estímulo ostensivo” por julgá-lo mais elegante.

A criação dessas fortes expectativas de relevância pelo estímulo ostensivo são consequências da própria definição de estímulo ostensivo e do Princípio Cognitivo da Relevância. Um estímulo ostensivo é projetado para atrair a atenção da audiência. Nos termos da TR, dada a tendência universal para maximizar a relevância, uma audiência prestará atenção somente a um estímulo que lhe pareça suficientemente relevante. Ao produzir um estímulo ostensivo, o comunicador conseqüentemente encoraja a audiência a presumir que esse estímulo seja suficientemente relevante para merecer ser processado. Mesmo um comunicador autocentrado, dissimulador ou incompetente manifestamente tenciona que a audiência assuma que o estímulo ostensivo é suficientemente relevante para ser processado. Esta é a base para o Segundo Princípio Comunicativo da Relevância, o qual é aplicado especificamente a uma comunicação inferencial ostensiva.

2.4.1.2.1 Princípio Comunicativo da Relevância

De acordo com a TR, todo estímulo comunicativo contém uma presunção da sua própria relevância ótima. O Princípio Comunicativo da Relevância e a noção de RELEVÂNCIA ÓTIMA são centrais para a teoria pragmática baseada em conceitos de relevância. Nesses termos, um estímulo ostensivo cria uma PRESUNÇÃO DE RELEVÂNCIA. A noção de relevância ótima é entendida como evidência do que a audiência de um ato de comunicação ostensiva está apta a esperar em termos de esforço e efeito.

a) Relevância Ótima

De acordo com a TR, um estímulo ostensivo é otimamente relevante se, e somente se:

- a. é relevante o suficiente para merecer esforço de processamento da audiência;
- b. é o mais relevante compatível com as habilidades e preferências do comunicador.

A compreensão do conceito de relevância ótima é importante para os objetivos desta pesquisa, como se mostrará nas recomendações sobre requisitos que um *website* deverá conter para servir a um usuário estrangeiro.

Para melhor aplicar esse conceito, podemos afirmar que a audiência pode esperar que o estímulo ostensivo seja, ao menos, relevante o suficiente para merecer ser processado. Dado o argumento de que um estímulo somente será merecedor de processamento se ele for mais relevante do que algum input alternativo disponível no momento, isso não é uma afirmação trivial. Na verdade, para satisfazer a presunção de relevância, a audiência pode ter de extrair uma conclusão mais forte do que, caso contrário, teria sido garantido. Em um exemplo apresentado por Sperber e Wilson em que um copo vazio é utilizado como estímulo ostensivo, se acontece de você somente notar meu copo vazio, você pode ser levado a concluir que eu poderia querer um drinque. Se eu deliberadamente balanço o copo para você, seria justificado concluir que eu gostaria de tomar um novo drinque.

A audiência de um estímulo ostensivo é levada a maiores expectativas, nos termos da cláusula (b) da definição de relevância ótima. Todos sabemos que um comunicador quer ser compreendido. É, conseqüentemente, de seu interesse – dentro dos limites de suas próprias capacidades e preferências – fazer com que seu estímulo ostensivo seja tão fácil quanto possível para a audiência compreendê-lo, e fornecer evidência não somente para efeitos cognitivos que ele objetiva alcançar, mas também para efeitos cognitivos posteriores que, por manter a atenção da audiência, ajudarão a alcançar sua meta. Sperber e Wilson (2002) assim exemplificam: o objetivo do comunicador poderia ser o de informar sua audiência que ele começou a escrever um artigo. O meio mais efetivo para a obtenção dessa meta poderia ser oferecer informação mais específica e dizer: “Eu já escrevi um terço do artigo!”. Nessas circunstâncias, sua audiência poderia, então, razoavelmente, entender que ele somente escreveu um terço do artigo, porque se ele tivesse escrito mais, dada a cláusula (b) da definição de relevância ótima, teria dito isso. A definição de relevância ótima postula que um estímulo ostensivo é aquele mais relevante que um comunicador está DISPOSTO e CAPAZ de produzir. O comunicador deverá, portanto, estar consciente para produzir um estímulo ostensivo que conduza a uma relevância ótima. Haverá portanto casos em que o

comunicador não quer ou não está em condições de transmitir algum estímulo ostensivo que comunicaria as suas intenções de forma mais econômica.

Ao analisar a contribuição para a relevância ótima, os autores Sperber e Wilson se dedicam também a apreciar a contribuição do silêncio, como elemento de informação. Quando uma resposta a uma questão for o silêncio, tanto pode ter ocorrido um silêncio intencional como uma situação em que o receptor da mensagem não é capaz de respondê-la. O silêncio, quando ostensivo, pode ser considerado um nível extra de intencionalidade, servindo conseqüentemente como via de COMUNICAÇÃO – ou de IMPLICAÇÃO – de que o destinatário da informação não está apto ou se recusa a responder. No quadro teórico de Grice (1975), no entanto, a vontade cooperativa do comunicador de prover qualquer informação requerida é tida como assegurada e por isso a noção de silêncios ostensivos e não ostensivos não é abordada. No contexto de Grice, a violação da primeira máxima de Quantidade já citada (“Faça sua contribuição tão informativa como requerida”) é atribuída à INCAPACIDADE do comunicador – em vez de uma RECUSA – de oferecer a informação requerida. A RECUSA de oferecer uma contribuição ‘tal como ‘requerida’ é uma violação do Princípio de Cooperação, e a suspensão do Princípio de Cooperação torna impossível transmitir qualquer implicatura conversacional, que é outro conceito elaborado pela TR. Para Sperber e Wilson, embora grande parte da comunicação seja cooperativa no sentido de que o comunicador está disposto a fornecer a informação requerida, a cooperação nesse sentido não é essencial para comunicação, como o é para Grice. Há, neste tópico uma divergência entre Sperber e Wilson e Grice, visto que para os primeiros a racionalidade em comunicação não requer cooperação no sentido griceano.

2.4.1.3 Relevância e compreensão

O Princípio Comunicativo da Relevância assegura motivação para o procedimento de compreensão exposto a seguir, princípio que a TR postula ser automaticamente aplicado ao processamento “on line” dos *inputs* esperados (pelo receptor ou destinatário). O ouvinte faz uso de uma estrutura conceitual construída por meio da decodificação linguística; adotando um caminho de esforço mínimo, enriquece esse caminho de forma explícita e implícita, até que o resultado

interpretado preencha as suas expectativas de relevância, momento em que o processo se encerra. Esse processo de enriquecimento é inferencial, visto que cada ouvinte é parte de um contexto específico.

Nos termos inferenciais da TR, além da lógica de decodificação abordada na teoria clássica da informação, o Princípio reconhece cada vez mais que mesmo o conteúdo explícito extrapola o que foi linguisticamente codificado.

Para sistematizar esse processo, a TR propõe a decomposição do processo em sub-tarefas, como se descreve a seguir no Quadro 3.

Quadro 3 - Sub-tarefas no processo global de compreensão

- a) Construção de uma hipótese apropriada sobre o conteúdo explícito (EXPLICATURAS) por meio da decodificação, desambiguação, resolução de referência e outros processos de enriquecimento pragmáticos.
- b) Construção de uma hipótese apropriada sobre suposições contextuais pretendidas (PREMISSAS IMPLICADAS).
- c) Construção de uma hipótese apropriada sobre implicações contextuais pretendidas (CONCLUSÕES IMPLICADAS).

Fonte: WILSON & SPERBER, 2005, p. 235

Nesta revisão conceitual, cabe ainda evidenciar o que se entende por “conteúdo explicitamente comunicado” ou explicatura, uma proposição recuperada pela combinação de decodificação e inferência, que fornece uma premissa para a derivação de implicações contextuais e outros efeitos cognitivos (SPERBER; WILSON, 1995). A noção de explicatura, considerada fundamental por Grice, é por ele dividida em duas categorias: a implicatura convencional e a implicatura conversacional. A implicatura convencional se materializa por meio de parâmetros lógicos. Por exemplo, “Ele é atleta, logo deve cuidar do seu físico” leva à implicatura convencional de que “todo atleta deve cuidar do seu físico”. Já as implicaturas conversacionais têm por função facilitar a compreensão na comunicação quando as implicaturas convencionais não operam ou não são suficientes para produzir eventos comunicativos satisfatórios. Tais implicaturas se realizam através do Princípio Cooperativo, regido por máximas que se encaixam em quatro categorias: Quantidade, Qualidade, Relação e Modo.

Diversos autores discutem, questionam e propõem aperfeiçoamentos aos estudos de Grice, entre os quais Searle (1969) e os já citados Sperber e Wilson (1995).

2.4.1.4 *Arquitetura Mental da TR*

As abordagens inferenciais da pragmática tratam a compreensão verbal como uma variedade da LEITURA DA MENTE ou TEORIA DA MENTE (a atribuição ou estados mentais para outros a fim de explicar e prever o seu comportamento). A ligação entre leitura da mente e comunicação é reconhecida pela abundância de evidências desenvolvimentais e neurológicas (SPERBER; WILSON; 2002). A literatura científica, no entanto, explica que a LEITURA DA MENTE em si tem sido analisada de diversas maneiras. É comum entre filósofos descrevê-la como um exercício de argumentação refletiva (um processo central de pensamento, conforme em Fodor (1986)). Sperber e Wilson declaram que em seus trabalhos iniciais também trataram a interpretação pragmática como um processo central e inferencial, espontâneo, intuitivo (e não consciente), um processo de reflexão. De acordo com esses autores, a tendência mais recente nas ciências cognitivas tem se orientado para uma crescente visão modular da mente ⁵. Concluiremos esta breve revisão teórica da TR explicando como poderíamos adaptá-la a perspectivas modulares de inferência e, particularmente, de leitura da mente.

A abordagem modular ou o uso de um mecanismo inferencial dedicado oferece a vantagem de permitir tirar proveito das regularidades de seu próprio domínio e conter procedimentos inferenciais específicos desse domínio. Na concepção modular de leitura da mente, Sperber e Wilson consideram que a psicologia tradicional de tipo “crença-desejo” pode ser substituída por procedimentos inferenciais de propósito especial, justificados por regularidades que existem somente nesse domínio. A literatura sobre leitura da mente explora exemplos de uso de um Detetor de Direção dos Olhos (*eye tracker*), que associa estados de percepção e de atenção com base nas fixações oculares e de um Detetor de

⁵ Conceito de módulo: um mecanismo computacional autônomo limitado a um domínio ou tarefa específica (SPERBER, 1996).

Intencionalidade capaz de interpretar movimentos autopropulsionados em termos de objetivos e desejos.

Outras considerações sobre a TR e arquitetura mental no campo da psicologia não seriam pertinentes aprofundar para os propósitos da presente tese, visto que o que se explora mais adiante será a combinação do conceito de estímulo ostensivo, oriundo da TR, com a multimodalidade no contexto desta pesquisa, nas considerações sobre a usabilidade de websites de bibliotecas digitais científicas.

2.5 Teoria da Relevância e Tradução

A teoria da relevância também tem sido estudada por alguns autores buscando seu uso em uma abordagem científica da tradução, com o objetivo de superar limitações de várias teorias que se fundamentavam sobre conceitos de fidelidade, equivalência ou funcionalidade. Trata-se de melhor estudar os processos de tradução à luz dos conceitos de uso descritivo e uso interpretativo. O uso descritivo de representação estabelece uma relação entre uma configuração mental (uma representação ou um conjunto de representações) e um estado de coisas tido como real no mundo físico ou num mundo fictício. Por sua vez, o uso interpretativo de representações é a relação entre duas representações mentais que apresentam propriedades lógicas e efeitos contextuais em comum.

Segundo a TR, a mente humana é capaz de manipular e operar com representações mentais através desses dois tipos de uso, segundo Gonçalves (2005). Em estudos de Gutt (1989), Alves (2005) e de Alves e Gonçalves (2005), foi desenvolvido o conceito de semelhança interpretativa entre enunciados como fundamento para se estudar o processo tradutório, a partir da noção de uso interpretativo das representações para comunicação ostensivo-inferencial, proposta pela TR.

No entanto, ainda não se conhecem estudos em que são utilizados esses conceitos de teoria da relevância e tradução em processos de tradução auxiliada por computador, um dos temas de interesse desta pesquisa com vistas à proposta metodológica da tese.

2.6 Multimodalidade

A expressão *multimodalidade* surgiu de uma busca de Kress e van Leeuwen (2006) por uma terminologia comum a todos os modos semióticos que são distintos e podem ser usados simultaneamente na comunicação de um significado. A linguística tradicional concebeu a linguagem como articulação de forma e significado, mas Kress e van Leeuwen (2001) assumiram que não era suficiente estudar um só modo de expressão, mas todos os modos possíveis por meio dos nossos sentidos. Assim, dedicaram-se ao estudo dos modos semióticos das diferentes linguagens e seus usos variados, inaugurando os debates sobre a multimodalidade.

Assim sendo, o termo multimodalidade surgiu para enfatizar a importância de se considerar outros aspectos semióticos além da linguagem em uso. Essa necessidade apareceu a partir do crescimento e da consolidação de outras formas de linguagem diferentes da escrita.

Portanto, a perspectiva multimodal leva em consideração que a língua em uso não ocorre por si própria, mas é integrada e dependente de outras formas de constituição e construção de significados (IEDEMA, 2003). Logo, os estudos multimodais têm o principal objetivo de observar esses processos de produção de significados situados socialmente e também oferecer meios para descrever uma prática ou representação em toda sua riqueza e complexidade semiótica.

Seguindo essa abordagem, a maior parte dos textos envolve um complexo relacionamento entre textos escritos, cores, imagens, elementos gráficos e sonoros, o enquadramento, a perspectiva da imagem, espaços entre imagem e texto verbal, escolhas lexicais, com predominância de um ou de outro modo, de acordo com a finalidade da comunicação, sendo, portanto, recursos semióticos importantes na construção de diferentes discursos que podem viabilizar a explicitação da informação.

A gênese do construto multimodalidade, conforme sinalizado anteriormente, está ligada à ênfase de aspectos semióticos. Neste sentido se faz necessário entender a semiologia que se originou de um trabalho de Ferdinand de Saussure que concebeu uma disciplina que estudaria o impacto dos signos no âmbito da sociedade, considerando os conhecimentos provenientes da Psicologia. Saussure

(1975) concebe a língua como “um sistema de signos que exprimem ideias, e é comparável, por isso, à escrita, ao alfabeto dos surdos-mudos, aos ritos simbólicos, às formas de polidez, aos sinais militares”, sendo o signo linguístico entendido como produto da associação significado/significante, ou seja, de um conceito e de uma imagem e ressaltando ainda que o significante é imotivado, isto é, arbitrário em relação ao significado, com o qual não possui nenhum laço natural na realidade.

Porém, é a partir de Barthes (1964) que os semiólogos passam a focar os signos de linguagens não-verbais com ênfase, produzindo trabalhos que abordaram a semiose humana em áreas como moda, comida, entretenimento, arte, música e fotografia. Todavia, Barthes ainda considera a linguagem escrita como o meio mais relevante para elaborar e articular outros modos de representação, sendo esta apresentada como principal e não subordinada a outras formas semióticas.

Considerando que a semiótica é a teoria geral dos modos de significar dentro do texto, Bakhtin (1981) deixa claro que onde não há texto não existe objeto de investigação e de pensamento. Percebe-se, então, que esta perspectiva limitadora, onde se consideram apenas as realizações escritas ou orais, não se podendo englobar imagens, sons e outros modos semióticos, demandou o surgimento de uma abordagem mais abrangente, tema amplamente abordado por Kress e Van Leeuwen (2001). Esses autores são categóricos ao afirmar que a língua escrita está menos no centro e é menos importante como meio de comunicação e que produtores de textos usam, cada vez mais, múltiplos modos de representação e comunicação.

A multimodalidade permite melhorar a expressividade comunicacional, afastando-se da tradicional comunicação grafocêntrica, por meio da ampliação de uso dos diversos modos semióticos para melhorar a compreensão de mensagens.

2.6.1 A CI, a Arquitetura da Informação e a Multimodalidade

Tratar de multimodalidade nos remete de imediato às suas principais abordagens e ao conceito que lhe dá origem: o conceito de modo. Autor de reconhecida notoriedade nesse campo, Gunther Kress (2009), define **modo** como “um recurso social e culturalmente estruturado para portar significado.”

Entre os recursos enumerados por Kress estão a imagem, a escrita, leiaute, música, gestos, discurso, imagens em movimento e trilhas sonoras. Fala-se ainda de

outros itens que são resultantes do trabalho e que teriam significado nos seus contextos, tais como o mobiliário, vestimentas e alimentos.

De acordo com Carey Jewitt (2009), as principais perspectivas para o estudo da multimodalidade são:

- a) Multimodalidade social semiótica;
- b) Discurso multimodal (uma abordagem com base em gramática sistêmico-funcional com raízes na linguística);
- c) Multimodalidade interacional.

Cada uma dessas abordagens apresenta particularidades de sua origem, como se descreve brevemente nesta tese para situar multimodalidade no contexto da ciência da informação.

Um conceito chave para diversos autores sobre multimodalidade focaliza a mudança com relação à língua como portadora principal do significado. Para esses autores, a representação e a comunicação sempre ocorrem através de uma multiplicidade de modos, ou seja, significados são construídos, distribuídos, recebidos, interpretados e ressignificados através de diversos modos de representação e comunicação, e não somente através da linguagem. Vale dizer que a hegemonia da linguagem é questionada, visto que para tais estudiosos “todas as interações são multimodais” e a multimodalidade “dissente da noção de que a linguagem sempre desempenha o papel principal na interação, sem deixar de reconhecer que frequentemente isso ocorre” (NORRIS, 2004, p. 2).

As abordagens citadas estão associadas ao trabalho de pesquisadores de origens distintas. A abordagem social semiótica, por exemplo, se apoia em grande medida nos trabalhos de Kress e Van Leeuwen (KRESS; VAN LEEUWEN, 2001; VAN LEEUWEN, 2005; VAN LEEUWEN, 2015). Como anteriormente mencionado, a abordagem com base em gramática sistêmico-funcional (SFG) se baseia por sua vez em trabalhos de Halliday e O’Halloran entre outros (HALLIDAY, 2004; O’HALLORAN, 2004). A multimodalidade interacional se fundamenta em trabalhos iniciais de Scollon e Scollon e Norris, dentre outros (SCOLLON; SCOLLON, 2003; NORRIS, 2004).

Embora não seja o propósito deste tópico explorar as abordagens da multimodalidade em maior profundidade *per se* — o nosso objetivo é analisar a sua

ligação com a ciência da informação — entendemos ser pertinente apresentar uma breve idéia do que significou cada uma delas como campo científico.

a) Multimodalidade social semiótica

O termo semiótica social foi introduzido pelo linguista Ballard e Michael Halliday, no título do seu livro *Language as Social Semiotics: the social interpretation of language and meaning* (1978). De acordo com este conceito, o “contexto” seleciona e organiza os recursos disponíveis para criar significado. Nesse trabalho, Ballard e Halliday questionam a separação tradicional entre língua e sociedade. Kress e Van Leeuwen (2001), por seu turno, estenderam a semiótica social, adaptando-a a um conjunto de modos.

b) Discurso Multimodal

Esta abordagem explora os conceitos de MDA – análise multimodal do discurso (*MDA, Multimodal Discourse Analysis*) – e SFG-MDA – análise sistêmico-funcional multimodal do discurso (*SFG-MDA, System Functional Grammar – Multimodal Discourse Analysis*). Vale, a esse propósito, mencionar que “discursos são conhecimentos socialmente construídos de algum aspecto da realidade ... desenvolvidos em um contexto social específico, sob formas que são apropriadas aos interesses dos atores sociais nesses contextos (...)” (KRESS; VAN LEEUWEN, 2001, p. 4). Para compreender as abordagens de MDA e SF-MDA, precisamos explicitar o significado de “recursos semióticos”, ou seja, ações, materiais e artefatos usados para comunicação entre pessoas. Van Leeuwen considera que a noção de recursos semióticos é originária dos trabalhos em linguística de Michael Halliday para quem a gramática de uma linguagem não é um código, nem um conjunto de regras para produzir sentenças corretas, mas um recurso para construção de significados. É ainda Halliday que identifica quatro tipos de potenciais significados, por ele denominados de metafunções da linguagem: a) significantes experimentais para construção de experiências; b) significantes lógicos para o estabelecimento de relações lógicas; c) significantes interpessoais para ação nas relações sociais; d) significante textual para organizar mensagens.

Introduzido por Michael O'Toole (1994) e, em 2006, por Gunther Kress and Theo van Leeuwen, a SF-MDA consiste em uma extensão da Teoria Sistemico-Funcional (SF) da linguística aos recursos semióticos, assim passando a incluir imagens visuais, simbolismo e imagens da matemática, música e som, movimento e gesticulação, arquitetura e espaço. Um exemplo ilustrativo de uma abordagem SF-MDA pode ser encontrado na análise de O'Toole sobre o Sidney Opera House, em que assinala:

Como uma cláusula numa linguagem, um edifício incorpora Tipos e Participantes de um Processo; suas funções específicas são modificadas em termos de material, tamanho, cor e textura; e seus elementos componentes são organizados de forma taxonômica como itens léxicos no vocabulário da nossa língua. (O'TOOLE, 2004, 15 *apud* JEWITT, 2009, p. 32)

c) Multimodalidade interacional

O conceito de multimodalidade interacional considera o discurso como uma das muitas ferramentas através das quais as pessoas agem. A análise multimodal interacional estuda uma forma de compreensão sobre como todos os objetos, todas as linguagens e todas as ações desenvolvidas com vários meios de mediação se interpenetram com um nexo de práticas sociais.

Na perspectiva da multimodalidade interacional, a atenção se desloca da representação e comunicação (foco dos trabalhos de Kress, Van Leeuwen e O'Halloran) para a **interação**. A multimodalidade expande o foco da interação explorando, além da linguagem, as formas como as pessoas empregam gestos, atenção, postura, movimento, espaço e objetos para mediar a interação em um dado contexto.

Uma das autoras que aborda a multimodalidade por esta ótica, Sigrid Norris, considera que:

Comunicação é uma interação se uma pessoa transmite uma mensagem e outra pessoa a percebe. Os modos utilizados para interagir não criam um momento comunicativo como interação, mas ao contrário o processo de fazer algo para ou com alguém permite-nos compreender um momento comunicativo como uma interação. (NORRIS, 2004, p. 149)

Para Norris, os modos não existem sem atores sociais que os utilizem de alguma forma. A análise interacional multimodal portanto se concentra no interrelacionamento entre modos num dado momento da interação social.

Na mesma perspectiva, Jones (2009, p. 114) argumenta que “não podemos presumir que uma configuração modal num certo momento utilizada por um ator social seja transportável para um momento diferente no tempo para ser utilizada pelo mesmo ou por outro ator”.

2.7 Tecnologias das Línguas

Como mencionado nos capítulos anteriores, o desenvolvimento acelerado da Internet conduziu a um processo de desterritorialização da língua e criou novas necessidades tanto no plano tecnológico como no seu uso. Em decorrência desse processo houve um crescimento exponencial da infraestrutura de informação e de inúmeras aplicações da super-rede, dentre as quais podemos citar: conexões em Banda Larga, dispositivos portáteis com acesso à Internet, acesso sem fio à Internet, dispositivos domésticos conectados, dispositivos multi-funcionais, conteúdo web customizado (para portáteis), acesso à Internet ilimitado, cartões pré-pagos, comércio eletrônico, M-comércio (E-comércio móvel), redes sociais e muitas outras aplicações e, naturalmente, diversas tecnogias das línguas.

Para os propósitos desta pesquisa, as tecnologias das línguas que se revelaram de maior interesse, como mostra a sua breve descrição apresentada a seguir, são: a) Verificação da linguagem; b) Busca na Web; c) Interação por Fala; d) Tradução Auxiliada por Computador ou Tradução Automática; e e) Sumarização automática (MENEZES, 2015).

a) Verificação da linguagem

Programas de verificação de linguagem, em geral, aplicam regras conhecidas aos textos. É o caso dos corretores ortográficos, de gramática ou de estilo embutidos na maioria dos processadores de texto.

Usuários de programas de edição de texto, como é o caso do MS Word e do BR Office estão habituados a utilizar ferramentas de correção ortográfica. Esses processadores de texto, em geral, por se tratarem de programas que não oferecem

uma revisão gramatical acurada, não oferecem a funcionalidade de correção gramatical. Há no entanto a oferta de diversas ferramentas de correção ortográfica do português, como é o caso do FLIP. Como corretor gramatical, podemos citar o CoGoOO. Temos ainda alguns programas de conversão do português ao acordo ortográfico firmado entre os países de língua portuguesa. É o caso do BRACorpus, de autoria do Prof. Marcos de Campos Carneiro, da UnB (2014).

A correção gramatical do português já contou com um *software* desenvolvido em colaboração da Universidade de Campinas (UNICAMP), com o Núcleo Interinstitucional de Linguística Computacional (NILC). Trata-se do ReGra (Revisor Gramatical), integrado ao MS Word e ao processador de texto REDATOR⁶, já desativado⁷.

b) Busca na Web

A geração atual de buscadores, em *intranets* ou em bibliotecas digitais, talvez seja a tecnologia de linguagem de maior uso. No entanto, para uma busca de informação mais elaborada, há necessidade de integrar conhecimentos linguísticos pouco utilizados nas arquiteturas de informação atuais. Além disso, usam-se também metadados, técnica usada para operacionalizar a recuperação de informações no campo da ciência da informação.

Algumas iniciativas em escala laboratorial já permitem a recuperação de informações com recurso a *thesauri* e bases de dados ontológicas. Alguns dos recursos já desenvolvidos para esse fim são as ontologias lexicais WordNet.PT e WordNet.BR⁸.

A expectativa no tocante à busca na Web será uma associação desses recursos aos conceitos da Web Semântica, ainda em fase incipiente de implementação prática.

⁶ Portal do Núcleo Interinstitucional de Linguística Computacional da Universidade de S. Paulo, Disponível em <http://www.nilc.icmc.usp.br/nilc/index.php/projetos>, Acesso em: 5 dez. 2015.

⁷ Notícia publicada na *Revista Inovação*, Ano 3, n. 1, jan./ fev. 2007 Instituto UNIEMP, <http://www.nilc.icmc.usp.br/nilc/projects/RevistaInovacao.pdf>

⁸ Essas WordNets são bases de dados de conhecimento linguístico do Português. Ver <http://www.linguateca.pt>, Acesso em: 30.mar.2017

A hegemonia do Google no momento é indiscutível, visto que na prática algumas iniciativas de motores de busca específicos para a língua portuguesa não conseguiram se consolidar. É o caso do AEIOU, em 1996 e do Sapo, em 1997 em Portugal e o Achei ou o Giga Busca, no Brasil.

c) Interação por Fala

A tecnologia da fala é a base para se criar interfaces que permitam ao usuário interagir com máquinas, por meio da voz em substituição ao “mouse”. Essa tecnologia se apoia em três dimensões principais:

- i. o reconhecimento automático da fala seleciona as palavras que foram pronunciadas em sequência pelo usuário;
- ii. A gestão do diálogo indica a ação a ser efetivada a partir da entrada vocal emitida pelo usuário e pela funcionalidade do sistema de interação por fala;
- iii. A síntese de voz (texto – para – fala) transforma a saída do sistema em sons para o usuário.

Conforme mencionado no Livro Branco sobre *A Língua Portuguesa na Era Digital* (BRANCO *et al*, 2012), os sistemas de reconhecimento de voz do português europeu e do português do Brasil têm apresentado bom desempenho.

O estado da arte da síntese de fala para o português (*output* vocal emitido pelo computador) é similar ao do reconhecimento de fala, apresentando desempenho satisfatório, e, por essa razão, algumas empresas estão atuando com sucesso nesse mercado. O exemplo mais flagrante de reconhecimento de voz é o SIRI, da Apple, que permite dar instruções vocais a um telefone celular. É o caso, também, das empresas Nuance (EUA, <http://www.nuance.com>), da SVOX (Suíça, <http://svoxmobilevoices.wordpress>) e da Voice Interaction (Portugal, <http://www.voiceinteraction.pt>). No mercado brasileiro, destaca-se a empresa DÍGITRO (<http://www.digitro.com>), que oferece produtos e serviços de fala-para-texto e de reconhecimento automático de fala, dentre outros.

Quanto à gestão de diálogo, a DigA – de domínio público – é a única aplicação completa para o português europeu. Tem-se notícia também da aplicação Olympus SDS, de código aberto, mas insuficientemente testada quando da publicação do Livro Branco citado.

Há ainda a registrar diversos *sites* que oferecem informações sobre ferramentas de tratamento automático da língua, tais como o LINGUATECA (<http://www.linguateca.pt>) e o NILC (<http://www.nilc.icmc.usp.br>).

d) Tradução Auxiliada por Computador (TAC), Tradução Automática (TA) ou Tradução por Máquina (TM)

A primeira experiência de TA real, do russo para o inglês, foi realizada em 1954, na Universidade de Georgetown, com um vocabulário reduzido (250 palavras), textos cuidadosamente selecionados e seis regras de sintaxe. Na época, essa experiência foi considerada satisfatória. Outro importante marco desse processo foram os estudos de Noam Chomsky (2014) sobre os conceitos de gramáticas formais, a partir de 1957. A União Europeia e países como África do Sul (11 línguas oficiais) e Índia (22 línguas oficiais) têm realizado programas de pesquisa nessa área. Iniciativas como a *Universal Networking Language* (UNL)⁹ também são relevantes no campo da tradução automática.

A partir dos anos 1980, houve uma grande retomada dos trabalhos de pesquisa e aplicações da TA. Resultante desse novo impulso, surgiram sistemas de tradução comercial, tais como o *Systran*. O *Eurotra*, um sistema europeu em constante desenvolvimento, também se iniciou nessa época.¹⁰

Entre os sistemas de TA que funcionam como auxiliares do tradutor humano, cabe mencionar: o *Trados Workbench* (<http://www.trados.com/>), o *IBM Translation Manager* (<http://www-4.ibm.com/software/ad/translat/>) e o *Déjavu* (<http://www.atril.com>). Há, ainda, sistemas que efetuam uma pré-edição do documento original, de forma a criar uma versão em linguagem mais simples, como a usada pela Xerox no *Systran* (<http://www.systransoft.com>) para traduzir seus manuais técnicos em várias línguas. Outros programas de tradução encontrados na literatura são: Google Translator, Worldlingo, Applied Language, Epals, Reverso, Prompt, Free Translation, Linguatec, Intertran e Moses.

⁹ Portal da Fundação UNDL, Disponível em <http://www.undl.org>. Acesso em: 30 abr. 2015.

¹⁰ Portal *Systran* (<http://www.systransoft.com>), *Eurotra* (<http://www.ccl.kuleuven.ac.be/about/EUROTRA.html>), Acesso em: 30 abr. 2015.

Cabe ainda dar destaque ao trabalho da Direção-Geral da Tradução da Comissão Europeia, que utiliza a tradução estatística por meio do *software* MOSES (<http://www.statmt.org/moses/>), o qual pode trabalhar com quaisquer pares de línguas, desde que sejam realizadas a preparação de corpora (“training pipeline”) e desenvolvido o decodificador (“decoder”) para cada língua. O modelo de tradução do MOSES se apoia nas técnicas de *phrase-based MT*, *syntactic MT* e *factored MT*, descritas na documentação desse sistema. O detalhamento sobre o MOSES está disponível no *link* já citado.

Nesse mesmo tópico, merecem ainda destaque as pesquisas de POTET (2009) cujo foco se concentra na otimização do texto traduzido a partir da uma “escolha ótima” das sentenças traduzidas por diversos motores de tradução. O artigo “*Méta-moteur de traduction automatique: proposition d’une métrique pour le classement de traduction*” descreve em detalhe o procedimento e os critérios utilizados no *software* de seleção das sentenças para compor o texto alvo otimizado.

No entanto, apesar dos avanços recentes, restam ainda diversos problemas a serem resolvidos no campo da TA, como constatam diversos autores. Na citada publicação do NILC, considera-se que “(...) restam importantes questões de cunho lingüístico a resolver (semântico e pragmático-discursivo, principalmente), tais como ambigüidades, referências anafóricas, etc. Como consequência, o desenvolvimento de sistemas completamente automatizados, que consideram questões linguísticas e extralinguísticas de forma profunda, principalmente em domínios abertos ou línguas naturais irrestritas, após mais de 50 anos de pesquisa, ainda é um desafio para a área de TA.”¹¹

e) Sumarização Automática

A sumarização sempre acompanhou a humanidade, mesmo antes da criação da imprensa. No entanto, na sociedade atual a necessidade de economizar tempo tem tornado a sumarização uma atividade mais premente, daí o surgimento da sumarização automática (sem contar a frequente sumarização de palavras, como ocorre nas comunicações por computador, particularmente, nas redes sociais).

¹¹ AMANCIO, D. R. E NUNES, M G. V., Avaliando Tradução Automática e Simplificação Textual em Redes Complexas, *Relatório Técnico NILC-TR-09-09*.

O indiscutível crescimento do volume de dados digitais na sociedade da informação já alcançou números que superam a casa dos *petabytes* (10^{15} bytes) e *exabytes* (10^{18} bytes) tornando difícil processá-los com o uso de ferramentas ou aplicações de processamento de dados tradicionais. Vive-se numa sociedade na qual há necessidade de localizar e processar informação da forma mais rápida possível e é nesse contexto que surgiram as técnicas de sumarização automática de textos, cujo início data da década de 1950 com os trabalhos de Luhn (1958), na empresa IBM.

Uma conceituação bastante simples formulada por Hovy (2009) define um sumário como um texto produzido a partir de um ou mais texto(s) cujo tamanho não pode ser superior à metade do tamanho do original. O conceito de texto inclui documentos multimedia, documentos *on-line*, hipertextos, etc. Diversos tipos de sumário são considerados na literatura técnica sobre o assunto: Indicativo, Informativo (Crítico), Extratos e *Abstracts*. Vale mencionar também que a sumarização está associada ao gênero (Manchetes (*headlines*), Breve descrição (*outlines*), Minutas (*minutes*), Biografias (*biographies*), Abreviações (*abridgments*), Resumo de Filmes (*movie summaries*), Cronologias (*chronologies*), entre outros. No âmbito desta tese, explora-se a viabilidade de sumarizar dissertações de mestrado (MENEZES, C.; BAPTISTA, D. M., 2017), como etapa de teste de uma metodologia para sua difusão entre estrangeiros proposta neste trabalho. De acordo com a literatura da área, sumários podem ser classificados como informativos, indicativos ou críticos. As abordagens de sumarização, usualmente consideradas em função da quantidade e do nível de conhecimento linguístico que utilizam, são denominadas abordagem superficial e abordagem profunda (em alguns casos pode haver mesclagem de técnicas dando origem a uma abordagem híbrida).

No entanto, a elaboração automatizada de sumários com a mesma qualidade de sumários elaborados por humanos depende de diversos fatores tais como comprimento e gênero do texto fonte, estilo de escrita e uso do léxico. A literatura técnica sobre o assunto traz diversos critérios para a escolha das sentenças que irão figurar no SA, tais como abordagens linguísticas, estatísticas e centradas na informação ou combinação de ambas. O sumário automático consiste na seleção de excertos do original portadores da maior ostensividade comunicacional, garantindo dessa forma a sua compreensão satisfatória por parte do usuário.

É também elemento importante para a sumarização automática o mapeamento de um determinado campo do conhecimento, fator que facilita a elaboração de sumários automáticos de boa qualidade. Um interessante exemplo de sumarização automática na área biomédica, o COMPENDIUM (LLORET *et al*, 2013), ilustra bem tal tecnologia e sua aplicação prática.

A técnica mais usual de montagem de sumários automáticos é a “sumarização por extração por varredura do texto” que se vale da criação de sumários mediante a seleção de excertos do documento original. No caso da sumarização humana, se requer a habilidade de entender, interpretar, criar um *abstract* e gerar um novo documento. Na SA, o procedimento é diferente: trata-se de classificar as sentenças do original de acordo com a sua relevância (ou com a sua ostensividade comunicativa) ou probabilidade de compor o “melhor sumário”, por meio de um ranqueamento baseado em métricas.

A sumarização automática em língua portuguesa tem recebido uma atenção especial de alguns grupos de pesquisa de universidades, principalmente em Portugal e no Brasil, como nas Universidades do Porto e de Lisboa, na UFSCar, na USP e na UFRGS, entre outras. Resultante desse trabalho, diversos sumarizadores de uso acadêmico estão em uso, como por exemplo o GistSumm, o NeuralSumm, o DMLSumm, o UNLSumm, o SuPor e o iSummarize.

No ambiente de *software* livre, a principal iniciativa, o GENSIM, permite a sumarização automática em diversas línguas.

A avaliação da qualidade de sumários também tem sido objeto de pesquisa e resultado em algumas metodologias de construção de métricas, como é o caso dos sistemas ROUGE e BLEU, para o cálculo de precisão, revocação ou cobertura e medida-F.

2.8 Considerações sobre a internacionalização de Bibliotecas Digitais

2.8.1 Contextualização

Como já foi mencionado no capítulo Introdução, as principais iniciativas para atender usuários falantes de diversas línguas no mundo digital apresentam como paradigma a elaboração de uma versão específica da biblioteca para cada idioma específico. É possível também encontrar muitas situações em que as páginas de abertura do *website* contemplam diversos idiomas, sem que possamos dizer que por isso a biblioteca passe a ser multilíngue. No caso brasileiro, essa situação também ocorre, ainda que nos faltem estatísticas mais específicas sobre o multilinguismo no mundo digital. A única pesquisa sobre a questão da internacionalização de *webpages* no Brasil, intitulada *Análise das características técnicas da Web brasileira: um estudo do .gov.br*¹², mostra que menos de 3% delas apresentam versão em outros idiomas, conforme mostrado na Tabela 1.

Tabela 1 - Distribuição dos principais idiomas das páginas Web no domínio .gov.br

Português		Inglês		Espanhol		Francês	
Número	%	Número	%	Número	%	Número	%
3.478.879	97,25	81.795	2,29	15.683	0,44	980	0,03

Fonte: *Dimensões e características da Web brasileira: um estudo do .gov.br*, 2011

Pode-se portanto afirmar que a temática da internacionalização do acesso a conteúdos em português disponíveis na *web* ainda não foi suficientemente desenvolvida no País. Algumas iniciativas em temas correlatos têm sido abordadas, principalmente nos núcleos de linguística computacional ou de processamento da linguagem natural de algumas universidades. Nesse aspecto, vale destacar o trabalho já mencionado do Núcleo Interinstitucional de Linguística Computacional (NILC), que desenvolve pesquisas nos seguintes temas: tradução automática, sumarização de texto, simplificação de textos, ferramentas de assistência à leitura e

¹² Pesquisa desenvolvida pelo W3C Brasil, Ceptro e CETIC e pelo NIC-BR, disponível em <http://ceweb.br/media/docs/publicacoes/4/cqibr-nicbr-w3c-censoweb-govbr-2011.pdf>. Acesso em: 17 ago. 2017

à redação, léxicos, dicionários, criação de *corpora*, linguística de *corpus*, ferramentas de análise sintática para diferentes níveis linguísticos (partes do discurso, sintaxe, semântica e discurso), avaliação automática, terminologia computacional, processamento da palavra e aplicações de técnicas de aprendizagem por máquinas para o processamento da linguagem natural, entre outros.

No tocante à oferta de serviços a estrangeiros, as unidades de informação ainda se encontram em estágio bastante rudimentar, não se tendo notícia do desenvolvimento de uma política para essa categoria de usuário.

A título de ilustração, observa-se que Cativelli (2009) realizou um interessante estudo em seu trabalho de conclusão de curso sobre os serviços oferecidos por bibliotecas universitárias modificados pelas tecnologias, como se mostra na Tabela 2.

Tabela 2 - Trabalhos nos SNBUs de 2000-2008 sobre informações nas bibliotecas universitárias brasileiras geradas pelas tecnologias: categorias de análise

Categorias / Anos	2000	2002	2004	2006	2008	TOTAL	%
Comutação		2	1	4		7	25
Biblioteca Digital		2		4	1	7	25
Serviço de Referência	1			3	2	6	21,42
Fontes de Informação	1	1	2	1		5	17,85
Automação		1	1		1	3	10,71
TOTAL	2	6	4	12	4	28	100

Fonte: Cativelli (2009, p. 29)

Vale realçar, no entanto, que embora tenha alcançado 21,42% do total de 28 serviços examinados os seis trabalhos presentes na categoria Serviços de Referência (SR) se distribuíram da seguinte forma: três enfocaram o SR na forma assíncrona, um trata do serviço de referência no geral, um aborda a utilização da ferramenta wiki e um trata da criação de um agente de conversação. Não havia portanto nenhuma categoria, atividade ou utilização de tecnologia orientada para usuários estrangeiros.

2.8.2 A Biblioteca Híbrida e os Serviços de Referência

O impacto das TIC nas unidades de informação tem sido muito expressivo e continua a ter seus reflexos na mudança de paradigma de funcionamento dessas unidades. Tomando como exemplo a biblioteca, novos serviços passaram a ser oferecidos e novas funcionalidades têm sido agregadas para melhor atender os usuários tanto presenciais como remotos. Sobre esses novos fenômenos, Levacov (2006) menciona que

A biblioteca, antes “ilha” de informações visitada fisicamente por usuários que buscavam um produto, principalmente o livro, precisa agora integrar-se e disputar espaço em um mar de informações. Seus usuários, tornando-se progressivamente remotos e descorporificados, acostumam-se a obter “aqui” e “agora” a informação digital “instantânea” disponibilizada pelas redes, em especial pela Internet e pela crescente largura de banda disponível. (LEVACOV, 2006, p. 210)

Dentro desse novo paradigma, uma das áreas amplamente afetada pelas novas técnicas são os serviços de referência, tradicional unidade de apoio aos usuários das bibliotecas.

Tal como a própria biblioteca, os serviços de referência também puderam se tornar híbridos, passando também a contemplar a dimensão digital na sua atividade. Sobre essa evolução, Pessoa e Cunha (2007) apresentam uma análise das perspectivas dos serviços de referência digital, tanto no que respeita à replicação a distância dos serviços anteriormente prestados e no que se refere à infraestrutura de informação e comunicação, como também sobre novos arranjos institucionais e cooperativos entre bibliotecas. Anteriormente, Márdero (2001) também abordara o tema, definindo os serviços de referência virtual e mencionando diversas redes e serviços de questões e respostas (Q&R) e softwares já lançados àquela época. Outros autores, como Dheeraj Sing Negi¹³, consideram haver dois tipos de serviços digitais de referências: *E-Mail Reference Service* e o *Real Time Digital Reference Service*. Para esse autor, os principais serviços de referência digital são:

¹³ Digital Reference Services in Digital Libraries, <https://www.slideshare.net/dheerajnegi9/digital-reference-service-36667646>, Acesso em: 15 jun. 2017

Collaborative Digital Reference Service (CDRS), Automatic Reference Librarians for the World Wide Web, Virtual Reference Desk (VRD)” e o *24/7 Reference*.

Por estar relacionado ao tema desta pesquisa, o serviço *Question Point*, hoje integrante do *Collaborative Digital Reference Service*, é baseado em uma rede de cooperação global e em ferramentas e tecnologias de comunicação, como citado por Pessoa e Cunha (2007). Uma particularidade interessante desse serviço consiste na manutenção de um perfil das bibliotecas, incluindo: escopo da coleção, horário de atendimento, especialistas nas equipes e idioma de cobertura do acervo. Esse perfil permite canalizar as perguntas para a biblioteca que estará mais apta a respondê-las. Esse, no entanto, foi o único serviço de referência virtual que encontramos nesta pesquisa que toma em conta a língua falada pelo usuário.

Pode-se, portanto, afirmar que, no tocante a serviços de referência que contemplam a dimensão do falante de outras línguas, há uma carência de literatura e de pesquisa sobre o tema e que o estado da arte ainda é incipiente nas bibliotecas de todo o mundo.

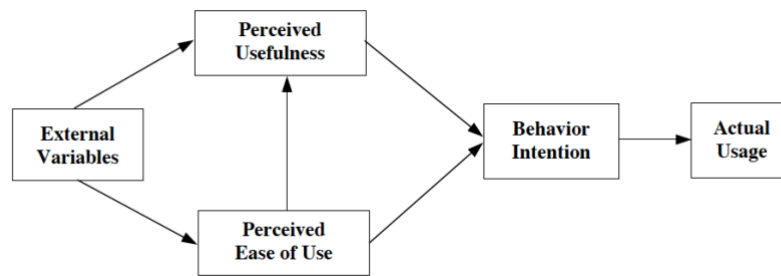
2.8.3 Usabilidade e Internacionalização de Bibliotecas Digitais

As bibliotecas digitais contemplam diversas dimensões: contextual, tecnológica, as dimensões de uso e as questões relacionadas à sua gestão (MARCONDES, C. *et al*, 2006).

No contexto desta tese, merece realce principal a dimensão da usabilidade da BD e mais particularmente as questões inerentes à interface do usuário.

Segundo o “Modelo de Adoção de Tecnologias” (TAM), de Thong, Hong e Tam (2002), os fatores que afetam a aceitação dos usuários são determinados pelas percepções de sua utilidade e facilidade de uso. O modelo proposto por esses autores contempla nove fatores agrupados em três dimensões, a saber: características da *interface*, contexto organizacional e diferenças individuais. A figura 3 ilustra o modelo proposto por esses autores para explicar a adoção de bibliotecas digitais por usuários.

Figura 3 - Modelo de Aceitação da Tecnologia de Thong, Hong e Tam.



Fonte: THONG et al. (2004)

No contexto desta tese, os três fatores inerentes às diferenças individuais tornam necessário que a *interface* da biblioteca digital contemple as especificidades de seus eventuais usuários estrangeiros.

3 METODOLOGIA DA PESQUISA

3.1 Elementos Conceituais

A investigação sobre modelos metodológicos para a elaboração de pesquisas tem sido objeto de trabalhos acadêmicos de diversas naturezas e aplicados às múltiplas áreas do conhecimento. No campo da pesquisa em Ciência da Informação, autores como Creswell (2010), Pickard (2013) e Costa (2014) oferecem diversos “modelos” adaptáveis a tais trabalhos de pesquisa.

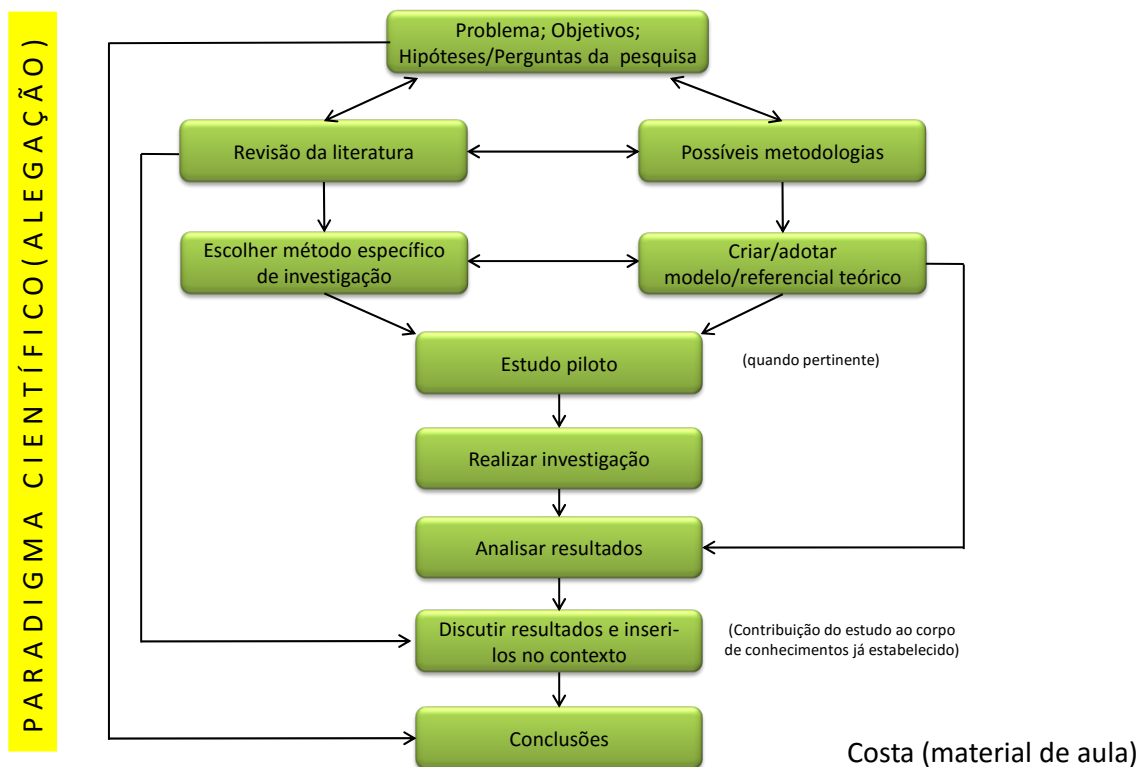
Creswell (2010), por exemplo, considera que três questões são basilares para um projeto de pesquisa, a saber:

- 1 Que alegações do conhecimento são feitas pelo pesquisador ?
- 2 Que estratégias de investigação vão orientar os procedimentos ?
- 3 Que métodos de coleta de dados serão usados ?

Pickard (2013) e Costa (2014) oferecem contribuições ao processo de pesquisa científica, indicando componentes indispensáveis a esse processo. Tais autores consideram que uma pesquisa acadêmica deverá estar composta por um paradigma, uma metodologia de pesquisa, métodos de pesquisa, técnicas de pesquisa e instrumentos de pesquisa, conformando dessa forma um conjunto integrado de elementos. Na figura 4 a seguir, Costa (2014) ilustra a concepção dessa autora sobre o processo de pesquisa científica:

Figura 4 - Diagrama conceitual de uma pesquisa científica

Processo da pesquisa científica



Fonte: Costa, Sely, Notas de Aula, Fundamentos da Ciência da Informação

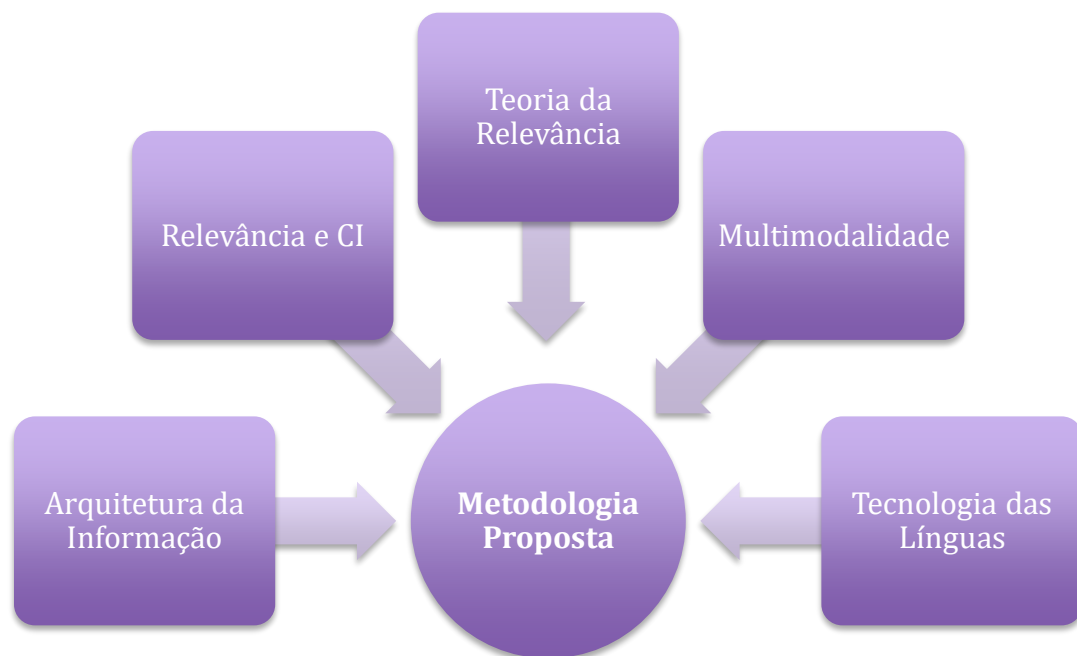
Com base em tais elementos conceituais, foi construído o arcabouço metodológico usado para implementação desta pesquisa, como descrito a seguir.

3.2 Descrição da metodologia da pesquisa

A proposta de tese está fundamentada conceitualmente no marco teórico mencionado nos capítulos anteriores (Arquitetura da Informação e Construção de Sentido; Relevância e Ciência da Informação; Teoria da Relevância: Multimodalidade e Tecnologias da Língua), como ilustra a figura 5. Trata-se, portanto, de uma pesquisa de natureza exploratória, interdisciplinar, destinada a facilitar o acesso e a compreensão de conteúdos científicos em português por estrangeiros, melhorando assim a vitalidade linguística da língua portuguesa no

mundo digital. A interdisciplinaridade da proposta de pesquisa envolve ciência da informação, linguística (pragmática), tradução automática e processamento da linguagem natural. Sua natureza exploratória reside tanto no que respeita a questões de expressividade comunicacional de *websites* (nos termos da TR), como predominantemente no tocante ao uso de tecnologias da língua como mediadoras e facilitadoras da compreensão do conteúdo em português por estrangeiros, através da oferta de um novo serviço a ser prestado por bibliotecas apoiados nos conceitos do Processamento da Língua Natural (PLN) e da Linguística Computacional (LC).

Figura 5 - Modelo Conceitual da Pesquisa



Fonte: Elaborado pelo autor

A identificação de demandas e especificidades de estudantes estrangeiros foi realizada por meio de questionários aplicados a falantes de francês, tanto no Brasil como na França, os quais não tinham proficiência em português, caracterizada pela aprovação no exame de conhecimento CELPE-BRAS. Esse exame avalia a compreensão e a produção oral e escrita da Língua Portuguesa de forma integrada.

As coleta de dados se desenvolveu da seguinte forma:

1. Pré-teste: Aplicação do questionário 1 (Anexo A) a estudantes de língua materna francesa, inscritos nos cursos de preparação do Núcleo de Ensino e Pesquisa de Português para Estrangeiros (NEPPE, <http://www.neppe.unb.br/br/>), da Universidade de Brasília, etapa realizada e concluída na fase inicial do projeto.
2. A partir dos resultados obtidos no pré-teste, elaboração de uma nova versão aperfeiçoada e ampliada do citado questionário para uso com estudantes franceses.
3. Aplicação do questionário 2 (Anexo C) a estudantes franceses, durante o período de doutorado sanduíche realizado no Grupo de Pesquisa GÉRIICO (Groupe d'Étude et de Recherche Interdisciplinaire en Information et Communication, <http://geriico.recherche.univ-lille3.fr/>), da "Université Charles de Gaulle" (Lille 3), na França.

Após a obtenção desses dados, sua tabulação e análise apresentadas no tópico seguinte, o trabalho de tese compreendeu o desenvolvimento dos seguintes componentes da metodologia proposta:

1. Uso de uma terminologia fundamentada na TR e na multimodalidade para descrever a metodologia proposta;
2. Identificação de componentes multimodais para recomendação no design de *websites*, de modo a facilitar seu uso por estrangeiros;
3. Seleção de tecnologias das línguas apropriadas para uso pelos especialistas das salas de referência virtuais. As tecnologias escolhidas foram a sumarização automática, os softwares de tradução auxiliada por computador e o alinhamento sentencial de conteúdos em línguas diferentes;
4. Integração dos componentes da metodologia em uma *suite* de programas de computador;
5. Descrição e documentação técnica dos serviços a serem ofertados ao usuário estrangeiro pelas salas de referência.

3.3 Implementação da Metodologia

A presente pesquisa conduziu à construção de uma metodologia de acesso por estrangeiros a conteúdos científicos em português. A metodologia faz uso de tecnologias da língua, por meio de um conjunto de ferramentas computacionais para facilitar a compreensão de conteúdos em português por estrangeiros.

No que respeita à usabilidade de *sites*, são feitas algumas recomendações para, através de estímulos ostensivos, aumentar a expressividade comunicacional de páginas *web* de bibliotecas digitais, objetivando atrair e manter a atenção do público alvo. A construção dos *websites* das bibliotecas digitais não faz parte da tese, embora também seja pertinente para o tema da internacionalização de bibliotecas e da arquitetura da informação para essa finalidade.

A implementação da metodologia compreende a definição de uma *suite* de programas para apoiar um serviço inicialmente oferecido pela sala de referência de bibliotecas, destinado a usuários estrangeiros e o protocolo de funcionamento de tal serviço. À medida que o serviço for implantado, é de se esperar que seus usuários irão adquirir autonomia no uso das ferramentas computacionais, por meio de orientação técnica oferecida pelas bibliotecas ou de auto-aprendizagem no uso do *software* em questão.

A interpretação dos questionários aplicados a falantes estrangeiros (sem proficiência na língua portuguesa com base no exame CELPE-BRAS) permitiu identificar algumas das necessidades dessa categoria de usuário quando se deparam com um conteúdo de uma biblioteca digital ou um repertório digital descrito em português.

A partir da tabulação e interpretação dos dados coletados, propõe-se uma solução técnica para atender a tais necessidades, por meio de um conjunto de serviços computacionais (remotos ou presenciais) oferecidos a serem prestados inicialmente pelas salas de referência e destinados especificamente a falantes de línguas estrangeiras. São também apresentadas recomendações de funcionalidades multimodais de arquitetura da informação a serem adotadas nos *websites* e nas páginas *web* de repertórios e bibliotecas digitais.

Vale observar no entanto e com ênfase que a arquitetura da informação dos *websites* deverá manter o seu objetivo principal, qual seja, o atendimento ao falante

do português. Deve-se evitar o risco de desvio da finalidade do website com um excesso de funcionalidades destinadas ao falante de outras línguas.

Quanto às tecnologias linguísticas, trata-se de — a exemplo de trabalhos já publicados no âmbito desta pesquisa (MENEZES; BATISTA, 2015; SANTOS, 2014) — utilizar técnicas tais como a conversão de arquivos (filtros e editores de textos), a formatação automática de textos, a sumarização automática, tradução auxiliada por computador e o alinhamento sentencial automático. É igualmente pertinente considerar a importância da difusão entre o público ao qual a metodologia se destina e os aspectos de custo dessa difusão em escala nacional e internacional.

O processo de integração entre as ferramentas de tratamento computacional da língua para apoiar os serviços a serem oferecidos pelas salas de referência constitui o componente sistêmico da metodologia proposta.

3.4 Coleta e Análise dos Dados

3.4.1 Coleta de Dados

A coleta de dados se realizou com dois grupos de usuários estrangeiros:

- 1) Onze estudantes de língua materna francesa inscritos nos cursos de preparação do Núcleo de Ensino e Pesquisa de Português para Estrangeiros (NEPPE, <http://www.neppe.unb.br/br/>), da Universidade de Brasília (Ver Questionário 1, Anexo A);
- 2) Quinze estudantes falantes de francês, inscritos no Mestrado (Master 1) da “Université Charles de Gaulle” (Lille 3), desenvolvido no Grupo de Pesquisa GéRiiCO (Groupe d’Étude et de Recherche Interdisciplinaire en Information et Communication, <http://geriico.recherche.univ-lille3.fr/>), na França (ver Questionário 2, Anexo C).

Houve também algumas tentativas de coleta de dados que se mostraram infrutíferas, devido a problemas de natureza distinta (tipo da pesquisa, pouco interesse de usuários por literatura científica em português, desistência na realização dos experimentos prescritos nos questionários, etc.). Tais experimentos indicam que uma coleta de dados de público estrangeiro baseada em experimentos envolvendo a língua portuguesa são mais eficazes quando feitas presencialmente.

O Questionário Nº 1 se compõe de quatro grupos de questões: 1) Identificação (6 questões); 2) Questões Gerais sobre o acesso dos entrevistados aos *sites* web na sua língua materna e em Português (7 questões); 3) Questões específicas sobre o website da Universidade de Brasília (8 questões); 4) Questões técnicas sobre os modos (*interfaces*) utilizados nos *sites* web pelos respondentes (11 questões). Teve como objetivo coletar informações com estudantes estrangeiros da Universidade de Brasília sobre as dificuldades encontradas no acesso à informação e na compreensão de seu significado em *websites* escritos em português, particularmente no ambiente da UnB.

O Questionário nº 2 está estruturado de forma análoga: 1) Identificação do respondente (9 questões); 2) Questões gerais concernentes ao acesso dos respondentes a *sites* web em francês e em português (10 questões); 3) Questões específicas relativas a *sites* web de conteúdo científico em português (21 questões); 4) Questões técnicas relativas aos modos (*interfaces*) dos *sites* web utilizados pelos entrevistados (11 questões). A exemplo do Questionário nº 1, teve como objetivo coletar informações sobre os principais obstáculos encontrados por estudantes de língua materna francesa no acesso e compreensão de textos em *sites* web escritos em português, particularmente no campo da literatura científica. Neste caso, foi formulado um número maior de questões com um nível maior de detalhe e procurou-se identificar o nível de uso de recursos linguísticos automatizados (tecnologias de informação e comunicação) pelos respondentes. Para esse fim, foram realizados dois experimentos de acesso a literatura científica, a saber: 1) acesso, por livre escolha, a uma tese de doutorado disponível na Biblioteca Digital Brasileira de Teses e Dissertações (<http://www.bdtb.ibict.br>); 2) acesso, por livre escolha, a um documento científico disponível em um dos repositórios seguintes: Diretório Luso-Brasileiro de Acesso Livre (Repositório de Acesso Livre OASIS (<http://oasisbr.ibict.br>, no Brasil) e Repositório Científico de Acesso Aberto de Portugal, RCAPP, (<https://www.rcaap.pt>, em Portugal), Biblioteca Digital de Monografias de Graduação da Universidade de Brasília (<http://bdm.unb.br>, UnB/RiUnB), Dissertações de Mestrado da Universidade de Brasília (http://repositorio.unb.br/?locale=pt_BR, UnB/RiUnB), Teses de Doutorado da Universidade de Brasília (http://repositorio.unb.br/?locale=pt_BR, UnB/RiUnB) e SCIELO (*Scientific Electronic Library On-Line*, <http://www.scielo.org>).

A tabulação dos questionários encontra-se nos Anexo D (p. 129) e Anexo F (p. 153), por se tratar de documentos escritos em francês, motivo pelo qual se julgou mais adequado apresentá-los em separado, evitando-se assim uma mesclagem acentuada de francês e português no texto da tese.

3.4.2 Coleta e Análise dos Dados

3.4.2.1 Questionário 1 – Identificação

Esta seção do questionário coletou os seguintes dados: Curso na UnB, Nacionalidade, Escolaridade e Data de Chegada ao Brasil.

As respostas sobre o “Curso na UnB” indicam uma predominância de estudantes dos cursos de Letras, havendo apenas dois estudantes de Engenharia (Produção e Elétrica). Quanto à nacionalidade, a predominância foi de estudantes africanos de países de língua francesa, salvo um estudante do Haiti. As idades variam de 20 a 27 anos e a escolaridade na maioria dos casos indica graduação incompleta. A data de chegada ao Brasil está entre 27/01 e 01/05/2015. O questionário foi respondido no 2º semestre de 2015.

3.4.2.2 Questionário 1 – Questões Gerais

As sete questões formuladas nesta seção tiveram por objetivo obter uma percepção preliminar do nível de uso de *websites* em francês e das dificuldades de acesso e compreensão de conteúdos em *sites* em português por estudantes francófonos do NEPPE.

A análise das respostas apontou para duas conclusões: 1) os estudantes de língua francesa consultam habitualmente *websites* na sua língua materna sem maiores dificuldades; 2) esses estudantes encontram dificuldades expressivas no tocante ao acesso e à compreensão de conteúdos, quando acessam *sites* em português.

As conclusões ora citadas resultam da análise das respostas, em que:

- a) todos afirmaram acessar *websites* escritos em língua francesa;
- b) ainda assim, dois respondentes afirmaram ter dificuldade de acessar as informações procuradas nesses *sites* web escritos em francês; os

mesmos estudantes observaram ter dificuldade na obtenção da informação procurada nesses *sites*;

- c) de forma surpreendente, dada a sua proficiência insuficiente na nossa língua, três estudantes declaram não ter dificuldade de obtenção de informações em *sites* escritos em português;
- d) a maioria dos entrevistados, em um percentual de 64% (sete das onze respostas à questão 5), no entanto, declara ter problemas de distinta natureza, com predominância para o conhecimento limitado do português e a necessidade de selecionar informação entre muitas respostas, elementos apontados como dificuldades para a compreensão desses conteúdos;
- e) salvo dois estudantes, a grande maioria, expressa por um percentual de 64% dos respondentes, indicou ter dificuldade de compreensão do conteúdo de *websites* em português; entre as principais razões por tal dificuldade foram citados os termos desconhecidos do vocabulário do estudante, problemas de interpretação de mapas, insuficiência de conhecimento gramatical e ortográfico, dentre outros.

3.4.2.3 Questionário 1 – Questões sobre o *site web* da Universidade de Brasília

As oito questões desta seção permitiram conhecer, de forma bastante preliminar, o grau de uso do *website* da UnB por estudantes recém-chegados ao País e à Universidade.

Uma análise das respostas permite afirmar que há um número ainda extremamente limitado de uso desse *website*. Ademais, o conhecimento das seções do *site web* da UnB pelos estudantes estrangeiros é quase nulo. A maioria dos respondentes não visitou nenhuma das seções presentes nesse *site web*. A maioria quase absoluta desconhece os repositórios institucionais disponibilizados pela Biblioteca Central.

As conclusões acima resultam de uma análise das respostas desse questionário, em que:

- a) quatro estudantes afirmam nunca ter acessado o *site web* da Universidade de Brasília;

- b) entre os sete estudantes que afirmam ter visitado essa página web, as seções que conhecem são variadas; não existe uma concentração de conhecimento na página da Biblioteca Central, visto que apenas um estudante citou essa seção;
- c) quanto às seções visitadas, não há uma uniformidade nas escolhas, visto que nenhuma seção foi escolhida por mais de um estudante; a seção da Biblioteca Central foi citada apenas em uma das respostas;
- d) não há conhecimento sobre os conteúdos oferecidos nos repositórios institucionais publicados pela Biblioteca Central da UnB; todas as respostas sobre esse tema afirmaram desconhecer tais repositórios e portanto não fazer uso dos seus conteúdos.

3.4.2.4 Questionário 1 – Questões Técnicas

As onze questões desta seção ofereceram uma idéia das tecnologias da língua eventualmente utilizadas pelos respondentes. Permitiram igualmente coletar informações sobre recursos presentes nos *websites* consultados na sua língua materna e sua comparação com funcionalidades (video, recursos multimodais, etc.) destinados a facilitar o seu uso por estrangeiros em *websites* em português.

As principais conclusões preliminares que foram obtidas nessa etapa da pesquisa foram as seguintes:

- a) os recursos oferecidos tanto nos *websites* em francês como em português são análogos;
- b) o uso de tecnologias da língua se resume a programas de tradução automática, com preponderância do *Google Translator* ou de dicionários on line;
- c) a tecnologia de sumarização automática não é do conhecimento de nenhum dos respondentes;
- d) o uso de recursos informatizados para compreensão de conteúdos em português pelos estudantes do NEPPE se limita ao *Google Translator* e a alguns dicionários informatizados. Uma única exceção foi mencionada: o uso dos aplicativos *conjugame.net.br* e *sinonimo.net.br*.

As conclusões acima são consequência de uma análise dos dados, em que:

- a) todos os respondentes, à exceção de um deles, alcançando um percentual de 91%, mencionaram a existência dos mesmos recursos e modos nos *websites* em francês e em português;
- b) embora todos afirmem fazer uso de recursos informatizados no tratamento da língua, as respostas são contraditórias; a rigor, o uso dessas tecnologias se limita ao Google Translator, embora quatro respondentes (36%) tenham mencionado também outros programas (dicionários automatizados, interfaces vocais, youtube), de forma pontual;
- c) no tocante à sumarização automática, houve quatro respostas afirmativas; no entanto, os respondentes ou não sabiam mencionar o nome do sumariador usado ou indicaram programas que não executam tal funcionalidade;

3.4.2.5 Questionário 2 – Identificação

Além dos dados pessoais, esta seção coletou informações sobre a nacionalidade, idade e número de anos de estudo da língua portuguesa.

As respostas mostraram estudantes de cinco nacionalidades diferentes, com predominância de idade entre 20 e 30 anos. Nenhum dos respondentes tinha estudado a língua portuguesa e todos estão inscritos no Mestrado em Ciência da Informação da Universidade de Lille 3.

3.4.2.6 Questionário 2 – Questões Gerais

As dez questões desta seção permitiram confirmar, como era esperado, o nível de exposição dos respondentes à língua portuguesa, o nível de acesso a *websites* em francês, o nível de sucesso na busca da informação procurada nesses *websites*, as razões do sucesso de tais buscas e fazer uma primeira identificação das razões citadas como obstáculo ao acesso e compreensão de textos científicos em português (obtidos nos *websites* especificados no experimento descrito no questionário).

As respostas permitem afirmar que os estudantes são frequentadores assíduos de páginas web em francês e que encontram dificuldades expressivas para acessar e compreender um texto científico em português. Essas conclusões se fundamentam nos seguintes elementos:

- a) todos os estudantes afirmaram utilizar com frequência *websites* em francês;
- b) nenhum dos estudantes é portador do diploma CELPE-BRAS, nem tem experiência no uso ou esteve exposto a um ambiente lusófono;
- c) 73,3% dos estudantes informaram que a existência de um menu para escolha da língua de uso facilita seu acesso ao conteúdo do website e 40% afirmaram que a existência de uma rubrica escrita em francês facilitou o acesso às informações procuradas;
- d) 66,7% manifestaram ter enfrentado dificuldades para encontrar um conteúdo nos *sites* lusófonos indicados no experimento de consulta;
- e) 93,3% expressaram ter dificuldade de compreender um documento recuperado nos *sites* indicados no experimento, sendo que 85,7% indicaram que a ausência de um menu para escolha da língua constitui um obstáculo importante para tal dificuldade.

3.4.2.7 Questionário 2 – Questões sobre recursos digitais de informação científica em Português

As quinze questões desta seção objetivaram conhecer o nível de compreensão e o grau de uso de tecnologias linguísticas utilizadas durante o experimento de consulta às bibliotecas digitais mencionadas no experimento.

De um modo geral, podemos concluir que o nível de compreensão foi mínimo e que o grau de conhecimento e uso das tecnologias linguísticas foi bastante insatisfatório.

Tais conclusões estão fundamentadas nos seguintes indicadores:

- a) Dos 46,7% de estudantes que consultaram dissertações de mestrado, 41,7% afirmaram não ter compreendido o seu conteúdo; vale realçar que entre os 58,3% que afirmaram compreender o conteúdo da dissertação, 55,6% indicaram ter compreendido apenas o título do documento, 33,3% afirmam ter compreendido o sentido de algumas

palavras e 11,1% entenderam somente alguns aspectos da dissertação; nenhum dos estudantes afirmou ter compreendido a totalidade da dissertação ou seu *abstract*;

- b) Dos 53,3% que consultaram uma tese de doutorado, 50% afirmaram compreender o seu conteúdo; no entanto, o nível de compreensão se revelou bastante limitado, visto que 33,3% manifestaram ter compreendido apenas o título da tese, 50% indicaram compreender o sentido de algumas palavras e 16,7% afirmaram entender somente alguns aspectos da tese;
- c) As respostas às outras questões desta seção confirmam que o nível de compreensão dos conteúdos científicos em português foi bastante limitado neste experimento.

3.4.2.8 Questionário 2 – Questões Técnicas

As onze questões desta seção se reportam ao nível tecnológico encontrado nos *sites* consultados (em francês e em português) e ao uso de tecnologias da língua para compreender o conteúdo científico dos *sites* visitados.

Duas conclusões principais podem ser obtidas das respostas: 1) os *websites* tanto em português como em francês oferecem recursos análogos, com predominância de textos explicativos; 2) o conhecimento e uso de tecnologias da língua para acessar e compreender conteúdos em português não alcança níveis expressivos. A fundamentação para tais conclusões encontra suporte nos seguintes dados coletados:

- a) a disponibilidade de textos explicativos em *sites* web alcança respectivamente 86,7% em francês e 83,3% em português; o modo imagem se situa em segundo lugar, com 73,3% em francês e 33,3% em português; a distribuição com relação aos outros modos é análoga, como mostram os gráficos do Anexo E;
- b) apenas um (6,7% das respostas) estudante afirmou ter usado um sumariador automático de textos em português e 26,7% dos estudantes afirmaram ter usado recursos informáticos para lhes auxiliar na compreensão de textos escritos em português;

- c) seis estudantes mencionaram o uso de programas de tradução automática do português para o francês; no entanto, apenas quatro mencionaram ter compreendido o texto traduzido; entre os que utilizaram programas de tradução, a avaliação da utilidade desse programa foi considerada insatisfatória, visto que nenhum atribuiu nota máxima à qualidade do texto traduzido.
- d) Quanto ao uso de programas de avaliação da qualidade da tradução, 93,3% não fazem uso de tais ferramentas.

4 ACESSO E COMPREENSÃO DE CONTEÚDOS EM PORTUGUÊS POR ESTRANGEIROS EM BIBLIOTECAS DIGITAIS CIENTÍFICAS: METODOLOGIA PROPOSTA

4.1 Contexto

A análise dos dados empíricos coletados junto a estudantes estrangeiros de língua francesa nos permitiu chegar a constatações importantes, a saber:

- a) o uso de tecnologias digitais e em particular das tecnologias digitais da linguagem na busca de informações científicas em língua estrangeira ainda se restringe a um número pouco significativo de estudantes (36% declararam usar outros programas de tratamento automático da língua, além do Google Translator, no pré-teste; no experimento realizado na França, 93,3% declararam nunca ter utilizado *softwares* de sumarização automática e 73,3% informaram nunca ter utilizado recursos informáticos para compreender o conteúdo de escritos em português);
- b) entre os estudantes estrangeiros que utilizam tais tecnologias o conhecimento da oferta de ferramentas se limita a programas de tradução por máquina (principalmente o Google Translator); o desconhecimento das tecnologias de sumarização automática e alinhamento sentencial e de outras ferramentas disponíveis para o tratamento automático da língua ainda predomina entre esse público;
- c) a ausência de uma oferta de serviços de apoio a usuários estrangeiros em *sites* web ou em bibliotecas dificulta o acesso dessa categoria de usuários à literatura científica em outras línguas e, em particular, em língua portuguesa.

4.2 Usabilidade e Internacionalização de Websites

No contexto da presente tese, uma questão pertinente que se coloca está relacionada aos requisitos para que um *website* se torne mais acessível ao usuário estrangeiro. A literatura aborda essa questão lançando a questão sobre “o que é um site internacional” e “o que é um *site* multilíngue”.

Para o *World Wide Web Consortium* (W3C), “Apesar de as palavras “internacional” e “multilíngüe”, às vezes, serem usadas uma pela outra, quando empregadas na descrição de *sites*, têm conotações bastante específicas, sendo assim primordial fazer uma clara distinção entre esses dois conceitos. Compreender bem tais diferenças é de suma importância na construção de *sites*, de maneira a melhor definir seus requisitos. Ao idealizar, desenvolver e estabelecer um *site*, surgem várias especificidades que vão demandar diferentes soluções técnicas”.¹⁴

A distinção conceitual entre um *site* internacional e um *site* multilíngüe determina e delimita questões bastante pertinentes para a metodologia aqui proposta.

De acordo com o W3C, essa distinção pode ser explicada da seguinte forma: “um *site* “internacional” é destinado a um público internacional, ao passo que um *site* “multilíngüe” usa vários idiomas. Um *site* internacional pode ou não ser multilíngüe, da mesma maneira que um *site* multilíngüe pode ou não ser internacional.”

No contexto desta pesquisa, podemos concluir que a maior relevância se situa na construção de um *site* internacional, que não será obrigatoriamente multilíngüe. A analogia com os *sites* acessíveis a pessoas com necessidades especiais é mais uma vez pertinente. Tais *sites* oferecem algumas funcionalidades para cegos, surdos-mudos ou portadores de necessidades especiais, mas continuam sendo *sites* monolíngües e grafocêntricos¹⁵. Na nossa proposta, os *websites* internacionais deveriam incluir funcionalidades para falantes de outras línguas, como se depreende das observações colhidas nos questionários aplicados.

A esse propósito, vale lembrar os dados já mencionados na seção 3.4.2.6 onde um expressivo número de respostas indica a necessidade de funcionalidades específicas de forma a personalizar em uma certa medida os *websites* com algumas funcionalidades para usuários estrangeiros, a exemplo do que se faz com a acessibilidade para cegos e surdos-mudos.

De certa forma, a coleta de dados permite confirmar nossa hipótese de que as funcionalidades encontradas por estrangeiros nos *websites* habitualmente por eles

¹⁴ Disponível em: <<https://www.w3.org/International/questions/qa-international-multilingual.pt>>. Consulta em: 26 maio 2017.

¹⁵ Sites com pouca presença da multimodalidade nos quais predomina o texto como modo de transmissão de mensagens

consultados na língua francesa seriam elementos facilitadores de uma navegação em webpages de bibliotecas digitais escritas em português, a saber:

- disponibilidade de um menu para a escolha do francês como língua do *site web*;
- existência de uma rubrica específica em francês para explicar o conteúdo do *site web*;
- página de abertura com ilustrações adequadas ao uso do *site web* por francófonos;
- descrição oral em francês;
- video detalhando as diferentes rubricas do *site web* e do seu conteúdo em francês;

4.3 Componentes da Proposta da Tese

A pesquisa realizada no âmbito desta tese para alcançar os objetivos gerais e específicos enumerados na seção 1.3 nos conduziu a uma análise de viabilidade para criação de um serviço personalizado e destinado aos usuários estrangeiros nas bibliotecas. Esse serviço estará apoiado em um dispositivo computacional (que poderá também ser utilizado por usuários remotos) composto dos seguintes componentes: a) Filtragem e Compatibilização de Formatos (FCF); b) Sumarização Automática (SA); Tradução por Máquina (TM); e d) Alinhamento Sentencial Automático (ASA).

Uma reflexão interessante sobre tais artefatos de mediação do acesso à informação nos é oferecida por Payeur (2013). Segundo essa autora,

por intermédio de artefatos mediadores se apresentando sob diferentes suportes (computadores, telefones móveis, tablets ...) e de diferentes ferramentas de produção, mas também de filtragem, de seleção, de informação (motores de busca, catálogos, ferramentas de seleção de informações segundo diferentes pontos de vista ...), o leitor, interagindo com outros atores, contrói percursos nos conteúdos digitais deixando rastros que podem ser ulteriormente reexplorados para fins de uma personalização da informação. (PAYEUR, 2013, p. 115, tradução do autor)

É, portanto, dentro do contexto de uma personalização da informação para o usuário estrangeiro que o dispositivo a ser descrito a seguir se coloca.

Um elemento importante na escolha do dispositivo computacional para o tratamento das tecnologias linguísticas usadas na tese é o ambiente computacional, neste caso o Python. Esta opção permite trabalhar com *software* livre e facilitará uma maior difusão e aperfeiçoamento da metodologia proposta, além de oferecer um custo mínimo para o usuário, visto que não haverá necessidade de pagar por licenças de *software* proprietário. Além disso, o ambiente escolhido deve oferecer recursos computacionais (tais como uma biblioteca de aplicativos) para implementação dos quatro componentes desse dispositivo. O ambiente Python preenche tais requisitos e permite uma difusão ampla da tecnologia para quaisquer pares de línguas. Além disso, como se percebe na implementação do dispositivo, há uma abertura para uso do português como língua fonte e diversas outras línguas alvo, elemento muito relevante para uma maior difusão de literatura científica em português no mundo digital.

Como se depreenderá da descrição a seguir, a partir de um texto obtido em uma biblioteca digital, o dispositivo computacional modular e iterativo proposto permitirá ao usuário dispor, no final do processamento, de um sumário extrativo automático do texto fonte em português associado ao texto na língua alvo por meio da tecnologia de alinhamento sentencial automático. O texto produzido em um formato aberto poderá ser tratado tanto por softwares livres como por *softwares* proprietários, para fins de adaptação a necessidades específicas de cada usuário.

4.4 Aplicativo Computacional para Facilitar a Compreensão do Texto da Biblioteca Digital Científica

Com o objetivo de permitir a disseminação e o aperfeiçoamento da metodologia proposta de forma ampla, a adoção de uma solução computacional baseada em *softwares* livres foi considerada a melhor opção para os propósitos da tese. As licenças de *software* livre permitem tanto o uso sem maiores restrições, assim como a modificação dos programas fonte, facilitando a melhoria do aplicativo por programadores de Python. A solução adotada, composta de quatro componentes integrados em um aplicativo disponível na web e desenvolvida em ambiente Python, é comentada a seguir.

a) Filtro e Compatibilizador de Formato (FCF)

Via de regra, além dos capítulos de maior interesse semântico sobre o tema abordado, os textos armazenados em bibliotecas digitais científicas (teses de doutorado, dissertações de mestrado, artigos científicos, relatórios técnicos) contém seções que não são relevantes para a elaboração de um sumário automático para compreensão do tema. A exemplo da nomenclatura usada em linguística para nomear *stopwords* e *stoplist*¹⁶, poderíamos denominar tais sessões informativas de *stopsessions*.

A título ilustrativo, o Quadro 4 mostra uma possível categorização das *stopsessions* para textos científicos depositados em repositórios digitais:

Quadro 4 - Lista de Stop Sessions

Capa	<p>É composta dos tópicos:</p> <ul style="list-style-type: none"> • Universidade • Departamento • Nível de graduação • Título • Nome do autor • Nome do orientador • Cidade • Data
Contra capa	<p>É composta dos tópicos:</p> <ul style="list-style-type: none"> • Universidade • Departamento • Nível de Graduação • Título • Nome do autor • Propósito do trabalho • Nome do orientador
Ficha Catalográfica	<p>É composta dos tópicos:</p> <ul style="list-style-type: none"> • Nome do autor • Título • Ano de publicação • Palavras-chave

¹⁶ Uma *stoplist* ou lista de termos proibidos é um dispositivo usado em indexação automática para filtrar palavras que tornaria pobre um índice de termos (FOX, 1989), tradução do autor

Banca Examinadora	Essa sessão é composta dos nomes dos examinadores da defesa e espaço para suas assinaturas.
Agradecimentos	Essa sessão é destinada ao autor do texto acadêmico, para que possa fazer os agradecimentos que acha necessário.
Resumo	Nessa sessão, o autor faz um breve resumo em sua língua materna sobre sua obra, e escolhe as palavras-chaves do texto.
Resumo em outros idiomas	Nessa sessão o autor traduz o resumo para outro idioma de sua preferência.
Índice geral	O índice geral é composto dos tópicos do texto e suas respectivas páginas.
Índice de figuras/tabelas	É composto da numeração das figuras/tabelas e suas respectivas páginas.
Referências Bibliográficas/Bibliografia	Sessão destinada às obras utilizadas no trabalho, ano de publicação e seus respectivos autores.

FONTE: Ribeiro (2016)

Outro item importante do dispositivo computacional são os formatos dos documentos digitais. Em geral, tais documentos são disponibilizados em formatos fechados, na maioria das vezes em pdf. Torna-se portanto necessário uma compatibilização desses formatos, a fim de que o arquivo de saída do filtro possa ser tratado automaticamente pelo sub-dispositivo seguinte, o sumarizador extrativo automático.

A busca de uma solução simples e de uso geral nos conduziu a escolher uma parametrização do texto pela simples indicação do intervalo das páginas a serem incluídas no arquivo a ser submetido ao sumarizador automático. Desse modo, o usuário do dispositivo para fazer a sumarização irá dispor de uma grande flexibilidade de escolha do texto que lhe convém sumarizar. Via de regra, pode-se recomendar o intervalo entre a primeira página do primeiro capítulo e a última página do último capítulo do texto a ser sumarizado.

b) Sumarizador Automático (SA)

O objetivo da sumarização textual automática está centrado na obtenção de um texto de tamanho mais curto capaz de representar o texto original mantendo a sua densidade semântica ou núcleo de relevância semântica, isto é, o sentido do texto original. Uma das formas de sumarização automática, denominada sumarização empírica, consiste na extração de sentenças relevantes do texto-fonte para a formação do sumário. Uma das principais vantagens desse critério é a simplicidade de geração do resultado, além do seu baixo custo. Objetiva-se garantir que o sumário gerado contenha (a) uma boa textualidade, ou seja, se suas sentenças são coesas e coerentes, e (b) uma boa proximidade, ou seja, se ocorre a preservação da idéia principal. Um exemplo de problema em (a) seria a seleção de uma sentença contendo um pronome, sendo que aquilo a que ele se refere está em sentença anterior e não selecionada (fenômeno conhecido como anáfora). Este é um problema de coesão textual típico na sumarização automática.

De acordo com Gupta e Lehal (2010), há diversas características a serem consideradas para inclusão de uma sentença em um sumário automático: palavra chave, palavra do título, localização da sentença, comprimento da sentença, presença de nomes próprios, palavras em caixa alta, frases introdutórias ou conclusivas, palavras escolhidas de um domínio do conhecimento, palavras destacadas pela sua fonte de caracteres, pronomes (desde que associados aos substantivos correspondentes), coesão entre sentenças (calculadas pela medida s de similaridade entre elas), coesão entre sentenças (calculada com base no centróide do documento), palavras de marcação do discurso (permitindo excluir do sumário automático sentenças que contenham tais palavras) e, finalmente, sentenças coerentes com uma análise do discurso (permitindo excluir sentenças periféricas à mensagem principal do texto).

Tais características são importantes e permitem implementar diversos métodos de sumarização textual automática extrativa, tais como: método TF-IDF (*Term Frequency-Inverse Document Frequency*); Método do agrupamento; Abordagem da Teoria dos Grafos; Abordagem da Aprendizagem por Máquina; Método LSA (*Latent Semantic Analysis*); Abordagem por obtenção de conceitos da

sumarização de texto (*Concept-obtained text summarization*); Sumarização de texto com redes neurais; Sumarização automática de texto baseada em lógica difusa (*fuzzy logic*); Sumarização automática usando regressão para estimar peso de atributos; Sumarização extrativa de múltiplos documentos; Sumarização textual extrativa baseada em perguntas e Sumarização textual extrativa multilingue.

Outro aspecto relevante para a sumarização automática está associado ao gênero do texto. No caso da literatura científica, há algumas experiências práticas de sumarização automática para um determinado gênero. No caso de *corpora* científicos, por exemplo, vale realçar o COMPENDIUM (LLORET *et al*, 2013), uma metodologia específica para sumarização de artigos da área biomédica. Outros sumarizadores automáticos citados por Lloret (2013) são o MEAD (para inglês e chinês), o MUSE (para inglês e hebreu), o SUMMA (com testes realizados em inglês, lituano, sueco e finlandês), o AZOM (para o persa) e o SummGRAPH (usado na sumarização automática em textos em espanhol de bio-medicina, jornalismo e turismo).

b.1) Sumarização automática em Português

A sumarização extrativa automática no Brasil se tornou um campo de pesquisa há pouco tempo, quando comparada com iniciativas análogas em outras línguas de larga difusão. Esse atraso pode ser atribuído, segundo alguns autores, à insuficiência de recursos específicos, tais como a disponibilidade de léxicos e de *corpora* que poderiam assegurar os fundamentos para abordagens profundas ou superficiais da SA (RINO *et al*, 2014).

Não obstante, em um nível ainda acadêmico, já estão disponíveis diversos sumarizadores automáticos do português, entre os quais destacamos o Gist Summarizer (GistSumm), o Term Frequency-Inverse Sentence Frequency-based Summarizer (TF-IDF-Summ), o Neural Summarizer (NeuralSumm), o Classification System (ClassSumm), o Text Summarization in Portuguese (SuPor), o ExtraWeb, o HTMLSUMM, o GEO e o SUMEX (SILVA, 2006). A escolha do “melhor sumarizador” da língua portuguesa ainda continua um tema em aberto, visto que diversas variáveis interferem na escolha, a principal delas sendo o gênero do *corpus*. Nesse particular, um trabalho interessante de avaliação dessas ferramentas foi realizado utilizando o TeMario, um *corpus* de 100 textos jornalísticos, acompanhados dos

respectivos sumários manuais e extratos ideais gerados automaticamente. Os sumarizadores avaliados foram o SuPor, o ClassSumm, TF-ISF-Summ, GistSumm, NeuralSum e algumas de suas variações (From-top e Random order) e foram calculadas medidas de Precisão, Revocação e a medida F¹⁷. Outro trabalho de avaliação da sumarização extrativa utilizou como *corpus* um conjunto de dezoito artigos científicos de computação retirados da *Revista Brasileira de Informática na Educação* e dos anais do *Simpósio Brasileiro de Informática na Educação – 1998* (SOUZA; NUNES, 2001). Esse trabalho avaliou quatro estratégias de seleção de sentenças a incluir no sumário automático: Palavras-chaves, Palavras-chaves + Localização, Palavras-chaves + Sinalizadoras e Palavras-chaves do Autor. Tais sumarizadores, no entanto, não têm sido usados para *corpora* de conteúdo científico, salvo em casos isolados com o sumarizador SUMEX (SILVA, 2016) acima citado e o GistSumm (BALAGE FILHO *et al*, 2007; MENEZES; BAPTISTA, 2017), ainda assim de forma muito limitada.

Além dessas questões, não há ainda uma comunidade de *software* livre muito ativa no campo do processamento da linguagem natural no Brasil e por isso os sumarizadores citados não atenderam aos critérios e objetivos desta pesquisa: integrá-los em uma metodologia de custo baixo e fácil difusão em nível internacional.

Por esse motivo, alargou-se a pesquisa para permitir a escolha de *software* livre pelas razões anteriormente anunciadas. No caso do sumarizador automático — o segundo componente do dispositivo proposto — a opção adotada foi o GENSIM,

¹⁷ As métricas usuais para avaliar a qualidade de sumários apresentadas pela literatura da área são a precisão, a revocação e a medida=F, assim definidas:

Precisão (P) = $\text{Correct} / (\text{Correct} + \text{Wrong})$, isto é, o número de sentenças do sumário automático presentes no sumário de referência em relação ao número de sentenças do sumário de referência (“baseline”).

Revocação ou Cobertura (C) = $\text{Correct} / (\text{Correct} + \text{Missed})$, isto é, o nº de sentenças do sumário automático presentes no sumário de referência (“baseline”) em relação ao número de sentenças do sumário automático.

Medida-F = $2 * (\text{Precisão} * \text{Revocação}) / (\text{Precisão} + \text{Revocação})$ (balanço métrico entre a Precisão e a Cobertura). (calculados a partir de um texto original de entrada e 2 extratos construídos pelo computador e por uma pessoa), onde:

Correct = nº de sentenças comuns aos 2 extratos.

Wrong = nº de sentenças do sumário automatizado ausentes do sumário elaborado pelo ser humano.

Missed = nº de sentenças extraídas pelo humano mas ausentes do sumário automatizado.

A medida-F indica a qualidade de cada sumário automático. À medida que se aproxima de 1, a medida-F evidencia que foram obtidos sumários de melhor qualidade e valores próximos de zero se referem a sumários de baixa qualidade em relação ao texto de referência.

cujo fundamento teórico se baseia no modelo TextRank (MIHALCEA; TARAU, 2004). Disponível em <https://github.com/RaRe-Technologies/gensim>, o GENSIM funciona em ambiente Python e requer a disponibilidade dos *softwares* científicos NumPy e Scipy¹⁸. Esse modelo permite construir um grafo representativo de um texto em linguagem natural e estabelece um ranqueamento entre os vértices que representam as sentenças do texto. Duas importantes aplicações do modelo TextRank são a extração de palavras-chave e a extração de sentenças para fins de sumarização automática (MIHALCEA; TARAU, 2004).

b.2) Uso do TextRank para Extração de Sentenças

Para fazer a extração das sentenças mais relevantes, o primeiro passo do algoritmo TextRank consiste na construção de um grafo associado ao texto. Como a unidade a ser ranqueada é a sentença, cada vértice do grafo representa uma sentença do texto a ser sumarizado.

Outro elemento importante deste método consiste no mapeamento das relações entre as sentenças. Esse mapeamento se desenvolve a partir do conceito de similaridade, uma medida do conteúdo de superposição de duas sentenças. Essa relação entre as sentenças pode ser interpretada como um processo de recomendação: uma sentença que aborda certos conceitos em um texto oferece ao leitor uma recomendação para referir-se a outras sentenças do mesmo texto que abordam os mesmos conceitos. Em consequência, um *link* pode ser associado entre essas duas sentenças que compartilham um conteúdo comum. A superposição (*overlap*) de duas sentenças pode ser determinada simplesmente como o número de *tokens* entre as representações léxicas das duas sentenças. Podem também ser usados filtros sintáticos para esse ranqueamento (por exemplo, substantivos e verbos, classes de palavras variáveis, etc.). Além disso, usa-se um fator de normalização para evitar a inclusão de sentenças muito longas no sumário. Formalmente, o conceito de similaridade entre duas sentenças pode ser definido conforme consta no quadro 5.

¹⁸ Detalhes sobre esses softwares estão disponíveis em <https://www.scipy.org/install.html>, Acesso em 14.set.2017

Quadro 5 - Definição de Similaridade

Dadas duas sentenças S_i e S_j , com uma sentença sendo representada pelo conjunto de N_i palavras que aparecem na sentença $S_i = w_1^i, w_2^i, w_3^i, \dots, w_{N_i}^i$, a similaridade entre S_i e S_j é definida como

$$\text{Similaridade}(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \& w_k \in S_j\}|}{\log(|S_i|) + \log(|S_j|)}$$

Fonte: Mihalcea e Tarau (2004, p. 7)

Obtem-se desse modo um grafo altamente conectado, com um peso associado a cada eixo, elemento que indica a similaridade entre as conexões estabelecidas entre os vários pares de sentenças do texto. O texto é representado por um grafo ponderado que serve para fazer o ranqueamento das sentenças com base no *score* de cada vértice. Esse *score* de cada vértice é calculado pela fórmula

$$S(V_i) = (1 - d) + d * \sum_{J \in I_n(V_\epsilon)} (1/| \text{Out}(V_j) |) S(V_j)$$

onde d é um fator entre 0 e 1 que tem o papel de integrar no modelo a probabilidade de saltar de um vértice para outro de forma aleatória. O grafo $G = (V, E)$ define-se como um grafo direcionado com um conjunto de vértices V e um conjunto de eixos E , onde E é um subconjunto do produto cartesiado $V \times V$. Para um dado vértice V_i , $I_n(V_i)$ é o conjunto de vértices que aponta para V_i (antecessores) e $\text{Out}(V_i)$ é o conjunto de vértices que emanam de V_i (sucessores).

Terminada a rodada desse algoritmo de construção do grafo, as sentenças são classificadas em ordem decrescente de *score* e aquelas mais bem situadas são selecionadas pelo programa de computador para inclusão no sumário.

Na versão implementada do TextRank, optamos por utilizar como parâmetro a taxa de compressão, isto é, a proporção do tamanho do sumário em relação ao texto original. Desse modo, as sentenças do grafo com maior pontuação serão escolhidas até alcançar a taxa de compressão fornecida como parâmetro do algoritmo.

c) Tradução Automática

A escolha do sub-dispositivo de tradução automática em ambiente Python recaiu no TextBlob (<https://github.com/sloria/TextBlob>), uma biblioteca para processamento de dados textuais. Além da tradução automática, essa biblioteca oferece quinze funcionalidades no tratamento automático da língua¹⁹, tais como: análise de sentimento, etiquetadores (“part of speech” *tagging*), *tokenization*, lematização, etc. A biblioteca TextBlob do Python trabalha com diversos pares de línguas, fator primordial na disseminação da metodologia proposta nesta tese. No aplicativo usado há disponibilidade para tradução do português para cinco línguas alvo: francês, inglês, alemão, espanhol e japonês.

O módulo de tradução da biblioteca TextBlob utiliza o *Google Cloud Translation API*, que faz uso do modelo *Neural Machine Translation (NMT)*. Embora haja poucos experimentos desse modelo usando Português como língua fonte, a tradução estatística (SMT = *Statistical Machine Translation*) tem se revelado um método promissor, de acordo com recentes experimentos (BENTIVOGLI *et al*, 2016; BOJAR *et al*, 2016; WU *et al*, 2016).

Na implementação do sub-dispositivo de tradução tornou-se necessário fazer a divisão do texto fonte, a fim de superar algumas limitações da versão do *software* livre adotada, na qual havia um teto de 1.000 palavras a serem traduzidas. Desse modo, foi elaborada uma versão específica para o módulo de tradução, sem limite do tamanho do texto fonte, o que permitiu usar a licença de *software* livre disponível em <https://github.com/sloria/TextBlob/blob/dev/LICENSE>.

d) Alinhamento sentencial automático

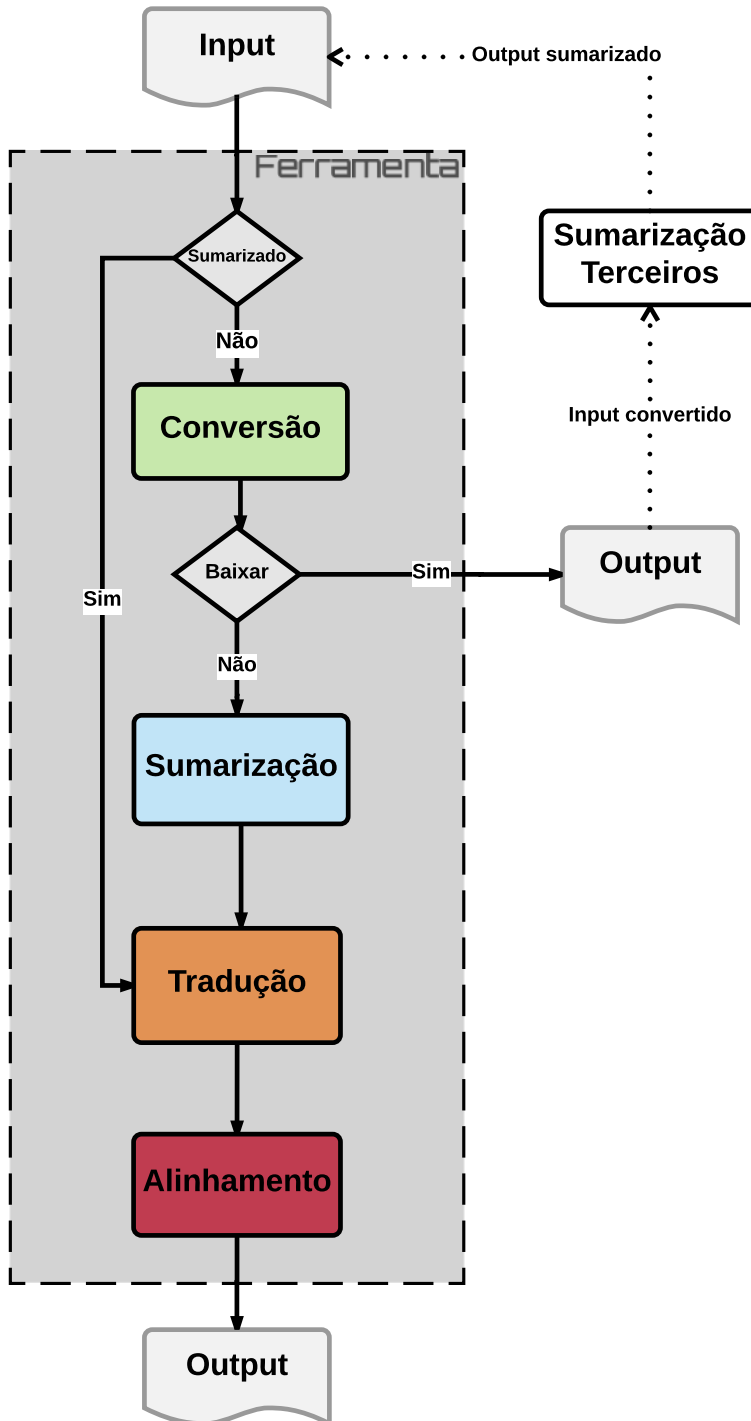
Última etapa do dispositivo, o alinhamento sentencial automático mostrará o resultado do aplicativo, exibindo um bi-texto com blocos de sentenças do texto-fonte em português e do texto traduzido para a língua alvo. Na escolha do sub-dispositivo para ambiente Python adotamos um API que implementa o algoritmo de alinhamento sentencial de Gale e Church (1993), atualizado em 10 de maio de 2012, disponível em: <https://github.com/vchahun/galechurch>.

¹⁹ Para uma descrição técnica dessas funcionalidades, ver <https://textblob.readthedocs.io/en/dev/>.

e) Fluxograma do Dispositivo

A figura 6 ilustra a integração de todos os componentes do dispositivo computacional desenvolvido para a implementação da metodologia proposta.

Figura 6 - Fluxograma do Dispositivo



Fonte: Elaborado pelo autor

f) Exemplos de Bi-textos

Para ilustrar o uso prático do aplicativo, utilizamos como *corpus* um subconjunto de 12 dissertações de mestrado do Programa de Estudos de Tradução (POSTRAD) da UnB.

Tais dissertações foram submetidas a dois experimentos e ajustes de formatação para inicialmente preparar o arquivo fonte. Desse modo, as dissertações foram sumarizadas tanto utilizando o GistSUMM como o GENSIM (um dos subdispositivos do aplicativo proposto na tese).

Em seguida, realizamos o processamento do dispositivo com os 24 sumários obtidos (com taxa de compressão de 20%) e tendo como resultado os bi-textos de saída produzidos pelo alinhamento sentencial automático. A figura 7 mostra uma página de texto de tese de doutorado²⁰, alinhado sentencialmente (convertida para HTML) e produzida pelo dispositivo. No Anexo H, mostram-se exemplos obtidos a partir das dissertações de mestrado do POSTRAD, usando dois grupos de sumários automáticos processados pelo GENSIM e pelo GistSUMM.

Figura 7 - Página de Bi-Texto com Alinhamento Sentencial

A complexidade deste problema pode variar dependendo se os genes foram abstraídos considerando a sua orientação, gerando permutações com sinal ou não. ||| La complexité de ce problème peut varier selon que les gènes ont été abstraits compte tenu de leur orientation, générant des permutations avec ou sans signe. ----- O problema de ordenação por reversões (usando permutações sem sinal) é um problema de otimização, onde o objetivo é minimizar o número de reversões para transformar um organismo em outro. ||| Le problème de la commande inverse (en utilisant des permutations non signées) est un problème d'optimisation, où l'objectif est de minimiser le nombre d'inversions pour transformer un organisme en un autre. -----
----- Os genomas dos organismos são representados como uma sequência de números naturais diferentes, onde cada número representa a ordem de um gene dentro

²⁰ O trecho usado foi obtido a partir da sumarização automática da tese de doutorado “Cálculo da Distância de Reversão e Construção de Árvores Filogenéticas usando a Ordem dos Genes”, de José Luís Soncco Álvares, depositada no Repositório de Teses e Dissertações da UnB, disponível em <http://repositorio.unb.br/handle/10482/23520>. Acesso em 29.ago.2017


do genoma. ||| Les génomes des organismes sont représentés comme une séquence de différents nombres naturels, où chaque nombre représente l'ordre d'un gène dans le génome. ----- A sequência mostrada representa uma solução ótima com só 4 reversões que é a distância de reversão entre estes dois organismos. ||| La séquence montrée représente une solution optimale avec seulement 4 inversions qui est la distance de réversion entre ces deux organismes. ----- Por outro lado, o rearranjo de genomas é baseado na análise das ordens dos genes, normalmente representadas como permutações com ou sem sinal. ||| D'autre part, le réarrangement des génomes repose sur l'analyse des ordres des gènes, généralement représentés comme des permutations avec ou sans signal. ----- Nesse contexto, a reconstrução de árvores filogenéticas está relacionada ao problema de rearranjo de múltiplos genomas, cujo objetivo é construir uma árvore filogenética que minimize o custo total da árvore com respeito a uma métrica (e.g. distância de reversão). ||| Dans ce contexte, la reconstruction des arbres phylogénétiques est liée au problème du réarrangement des génomes multiples, dont l'objectif est de construire un arbre phylogénétique qui minimise le coût total de l'arbre par rapport à une métrique (par exemple, la distance d'inversion). -----

Fonte: Elaborada pelo autor

g) Formatação do Texto Final

Como se constata dos exemplos já apresentados, os bi-textos podem ser utilizados por qualquer usuário ou biblioteca, visto que não estão identificados ou dependentes do seu ambiente computacional e utilizam um formato de uso bastante difundido, o formato **.txt**. É, no entanto, de bom alvitre que um serviço oferecido por uma biblioteca contenha identificadores da entidade prestadora desse serviço. Desse modo, uma etapa necessária nesse trabalho de acabamento do texto consiste na melhoria da legibilidade do bi-texto e da identificação da instituição ofertante do serviço. A título exemplificativo, o trecho do documento a seguir (Figura 8) mostra um bi-texto identificado com os dados da Biblioteca Central da Universidade de Brasília (UnB).

Figura 8 - Exemplo de Bi-Texto com Identificação da Biblioteca Central da UnB

 Universidade de Brasília Biblioteca central Serviço de acesso a conteúdos digitais em português para estudantes estrangeiros	
Título do Documento Original: Norma linguística e oralidade fingida na tradução de Persépolis	
Tipo do Documento: Tese de Mestrado	
Língua do Original: Português	
Língua Alvo: Francês	
Nome do Arquivo: Tese-pos-defesa-5dez.docx	
Sentença Fonte	Sentença Alvo
De certa forma, há uma ampliação dos recursos que serão aplicados na tradução: o autor utilizou a imagem que deve ser considerada pelo tradutor como chave de leitura	D'une certaine façon, il y a une augmentation des ressources qui seront appliquées dans la traduction: l'auteur a utilisé l'image qui doit être considérée par le traducteur comme une clé de lecture.
A partir dos elementos presentes na linguagem dos quadrinhos, o trabalho do tradutor dessas histórias não se restringe ao texto escrito é preciso levar em conta outros desafios que vão desde uma linguagem icônica até as várias normas linguísticas incidentes no texto.	À partir des éléments présents dans la langue de la bande dessinée, le travail du traducteur de ces histoires ne se limite pas au texte écrit, il faut tenir compte d'autres défis allant d'un langage iconique aux différentes normes linguistiques incidentes dans le texte.
No gênero quadrinhos, os balões expressam a fala de cada personagem e os autores usam os recursos gráficos para expressar a oralidade na escrita, o que leva o pesquisador (e, por conseguinte, o tradutor) a rever os conceitos de língua falada e língua escrita bem como suas inter-relações	Dans le genre comique, les ballons expriment le discours de chaque personnage et les auteurs utilisent les ressources graphiques pour exprimer l'oralité par écrit, ce qui conduit le chercheur (et donc le traducteur) à examiner les concepts de langue parlée et de langue écrite ainsi que leurs interrelations

Fonte: Elaborada pelo autor

4.5 Disseminação da Metodologia por Bibliotecas

A coleta de dados desta pesquisa constatou que, sem sombra de dúvida, há um elevado risco de desistência de consultas por usuários estrangeiros, ao se

deparar com informação científica disponibilizada em uma biblioteca digital de conteúdos em português. Esse comportamento pode ser explicado a partir de conceitos de relevância, visto que um texto em língua portuguesa que não pode ser compreendido oferece ostensividade comunicativa insuficiente para manter o interesse desse tipo de usuário. Essa atitude de desistência também se explica pelas práticas habituais de pesquisa, em que a reação natural conduz o usuário a buscar informação nas línguas que conhece, abdicando em alguns casos de utilizar literatura mais pertinente ao seu objetivo. Como já se mencionou anteriormente, trata-se de um processo comparável ao que ocorre com portadores de deficiências físicas (nesta caso alguns *websites* já oferecem ferramentas de acessibilidade bastante eficientes e eficazes). Infelizmente, quando se trata de um obstáculo cognitivo para compreensão de conteúdos digitais, como ocorre com o objeto de estudo desta pesquisa, a disponibilidade de ferramentas nos *websites* ainda é limitada (ou quase inexistente), fator que remete à necessidade de recorrer às bibliotecas como mediador fundamental na disseminação da metodologia ora desenvolvida.

Isto posto, a nossa proposta para disseminação da metodologia desenvolvida nesta tese se concentra inicialmente na oferta de um serviço especializado para usuários estrangeiros pelas bibliotecas.

Há duas principais razões para essa proposta de estratégia de difusão:

- a) como evidenciado no decorrer da pesquisa, os *websites* das bibliotecas digitais científicas não contemplam ainda especificidades inerentes ao usuário que não compreende português, particularmente no tocante às interfaces e a ferramentas de auxílio a essa categoria de usuário;
- b) em um primeiro momento, mercê da complexidade do assunto, o uso da metodologia e da ferramenta computacional ora proposta requer um grau mínimo de conhecimento e de treinamento do usuário, tarefa que a nosso ver indica a conveniência de intervenção técnica de pessoal especializado da biblioteca.

Para auxiliar o processo de popularização da metodologia, elaborou-se um tutorial (Anexo G, p. 180) com todos os detalhes e especificações técnicas necessárias ao seu uso. Por se tratar de um serviço oferecido via web, as bibliotecas

poderão prestar um serviço especializado a estrangeiros ou oferecer treinamento no uso desta metodologia de forma presencial ou à distância a quem queira se tornar autônomo no uso da *suite* de programas computacionais .

O desenvolvimento de um aplicativo em *software* livre também constitui fator que certamente irá facilitar a difusão e o aperfeiçoamento da metodologia e do *software* desenvolvido nesta tese. Vale ainda salientar a grande flexibilidade oferecida para a personalização dos documentos pelas bibliotecas que resolverem implantar a metodologia ora proposta.

4.6 Uso Remoto da Metodologia

Como se constata, a presente tese teve como objetivo ampliar a vitalidade da língua portuguesa no campo científico, por meio de uma metodologia inovadora, apoiada em uma aplicação computacional disponível na Internet. Por esse motivo, todo e qualquer usuário que se interesse por literatura científica em português, esteja onde estiver, poderá fazer uso da suite de programas construída e disponível na url: <http://164.41.62.101/>. Para tal, será suficiente familiarizar-se com as especificações do sistema, disponíveis no tutorial já mencionado, ou seguir um treinamento especializado que seja oferecido por bibliotecas que implantarem o sistema computacional citado.

5 BIBLIOTECAS DIGITAIS E REPOSITÓRIOS PARA APLICAÇÃO PRÁTICA DA METODOLOGIA

O campo de aplicação da metodologia proposta nesta tese compreende todo o amplo espectro de literatura científica digitalizada escrita em português, disponível em inúmeras bibliotecas digitais em todo o mundo. Para ilustração desse uso, foram realizados diversos experimentos práticos, a saber:

- a) a partir de um *corpus* de dissertações de mestrado do Programa de Pós-Graduação em Estudos da Tradução (POSTRAD) da Universidade de Brasília, efetivou-se o processamento computacional da *suite* de programas da tese, tanto a partir dos textos em pdf (sumarizados através do GENSIM) como a partir de sumários textuais automáticos realizados pelo GistSUMM (Anexo H).
- b) a partir de escolhas aleatórias de teses de doutorado realizadas em bibliotecas digitais científicas publicadas em português (Anexo I), efetuou-se esse mesmo procedimento computacional, até a obtenção dos bi-textos em francês e português, alinhados sentencialmente.

Os resultados obtidos, mostrados nos anexos, evidenciaram que a metodologia está plenamente operacional para uso o mais amplo possível, tanto através de serviço oferecido por bibliotecas como diretamente por usuários remotos devidamente capacitados.

Algumas das principais bibliotecas e repositórios digitais onde se pode encontrar literatura científica em português e aplicar os resultados desta tese são: Biblioteca Digital de Teses e Dissertações (BDTD, <http://www.bdttd.ibict.br/>); Scientific Electronic Library On-Line (<http://www.scielo.org>); Repositórios Internacionais de Literatura Científica em Português (RCAPP, <https://www.rcaap.pt/>); Repositórios Nacionais de Literatura Científica em Português (e.g. OASIS, <http://oasisbr.ibict.br/vufind/>) Repositórios de Acesso Aberto de Literatura Científica em Português; Biblioteca de Teses e Dissertações da Universidade de São Paulo (SIBI, <http://www.sibi.usp.br/>); Networked Digital Library of Thesis and Dissertations (NDTLD, <http://www.ndltd.org/>), Rede Pró-Intercom

(<http://www.portalintercom.org.br/a-intercom>), University of Macau Thesis Collection (UM E-Thesis Collection, http://libdigital.umac.mo/was5/um_theses/main.jsp) e inúmeros repositórios de teses e dissertações de outras universidades no País e no exterior.

Há que se notar que alguns dos repositórios de teses e dissertações protegem o seu acervo digital e requerem senhas de acesso para facultar o acesso e tratamento de seus documentos. Nesse caso, por óbvio, será necessário um entendimento prévio do usuário com os gestores do acervo de modo a liberar o acesso aos textos a fim de que o aplicativo desta tese possa ser processado com êxito.

Cabe ainda realçar que outras fontes de literatura científica, tais como portais de periódicos (como o Portal de Periódicos da CAPES, <http://www.periodicos.capes.gov.br>), oferecem conteúdos digitais que também podem ser processados pelo aplicativo desenvolvido nesta tese.

6 CONSIDERAÇÕES FINAIS

A vitalidade linguística no mundo digital tem se tornado um tema recorrente tanto nas organizações internacionais como nas comunidades linguísticas internacionais e nacionais. No âmbito do Sistema da Nações Unidas, a UNESCO há muito debate esse tema e, em sua Conferência Geral de 2003, seus estados membros aprovaram o documento “Recomendação da UNESCO sobre a Promoção do Uso do Multilinguismo e Acesso Universal no Ciberespaço” onde está evidenciada com destaque a necessidade de ampliar o acesso à informação no mundo digital nas línguas de todo o mundo. A temática continua na pauta da UNESCO e, em 2015, aquele órgão das Nações Unidas publicou o relatório “*A Decade of Promoting Multilingualism in Cyberspace*” (UNESCO, 2015), no qual são apresentadas ações relevantes empreendidas em favor do desenvolvimento de um ciberespaço multilíngue.

Na esfera dos países lusófonos, a Comunidade dos Países de Língua Portuguesa (CPLP) também tem realçado a importância de assegurar maior vitalidade da língua portuguesa no mundo digital, através de declarações de seus dirigentes e de estudos e eventos técnicos promovidos principalmente pelo Instituto Internacional da Língua Portuguesa (IILP). Nesse particular, a “Carta de Guaramiranga” (2012) enfatiza a importância da criação de novos recursos linguísticos e de meios digitais de difusão do conhecimento para permitir a instituição de um modelo aperfeiçoado de gestão das políticas da língua portuguesa, comum e mais representativo em todos os países.

Outro campo que guarda sintonia com a temática desta tese é o acesso livre a publicações científicas financiadas com recursos públicos. Ao facultar acesso livre por estrangeiros à literatura científica produzida em português, a metodologia ora proposta permite ampliar a difusão da ciência para outras comunidades linguísticas, em uma esfera mundial.

Nesta pesquisa acadêmica, ao estudar o acesso e compreensão de estrangeiros à literatura científica em português, como proposto nos objetivos gerais e específicos desta pesquisa, introduziu-se no campo da ciência da informação um fenômeno ainda pouco estudado e decorrente da instauração da sociedade da informação e do conhecimento.

O acesso à informação científica por estrangeiros, tema desta tese, depende de dois grandes componentes: 1) a interface do usuário nas páginas web das bibliotecas digitais e 2) os aspectos cognitivos que lhe permitam compreender os conteúdos dos documentos dessas bibliotecas.

Tendo em vista os objetivos fixados no início desta pesquisa, a metodologia ora desenvolvida permitiu demonstrar de forma simples e prática a viabilidade do segundo componente do tema, a fácil compreensão de literatura científica em português por estrangeiros por meio da utilização do aplicativo desenvolvido e da estratégia proposta para sua disseminação por bibliotecas.. Cabe também evidenciar que, embora os *corpora* utilizados sejam constituídos por dissertações de mestrado e teses de doutorado, o alcance da metodologia vai além desse tipo de literatura científica, visto que cabe ao usuário escolher o conjunto de páginas a ser submetido ao aplicativo que acompanha esta tese.

No decorrer deste trabalho de pesquisa, desenvolvido com o par linguístico Português – Francês, ficou evidenciado também que a metodologia parece poder ser extensível a outros pares de línguas, inclusive as que não utilizam o sistema de escrita latino (como é o caso do japonês). Seria portanto interessante dar prosseguimento a pesquisas dessa natureza, que irão contribuir para a ampliação da diversidade linguística no campo científico.

Outra fonte de pesquisa bastante promissora, compreende-se, são os elementos constitutivos da metodologia (sumarização, tradução automática e alinhamento sentencial automático). À medida que o estado da arte de cada um desses componentes evoluir, será apropriado avançar em pesquisas que irão aperfeiçoar os resultados desta tese. Um processo de vigilância tecnológica, portanto, se impõe com o objetivo de manter atualizada a metodologia aqui construída e sua difusão no mundo digital. É igualmente pertinente o controle de qualidade dos textos produzidos pela metodologia, através de métricas de avaliação da qualidade da sumarização, da tradução automática e do alinhamento sentencial automático. De particular interesse, neste caso, são as metodologias de avaliação de qualidade dos resultados de cada dispositivo e o uso de programas computacionais de cálculo das métricas de avaliação da qualidade dos textos sumarizados, traduzidos ou alinhados sentencialmente.

Quanto as interfaces, há uma ampla gama de possibilidades de pesquisas para fazer avançar o estado da arte. Neste campo situa-se principalmente a questão da internacionalização dos *websites* e do uso de arquiteturas de informação multimodais para torná-los acessíveis a estrangeiros, como se constata a partir dos dados levantados na presente pesquisa. Nesse particular, muito poderá ser aproveitado das pesquisas já realizadas sobre acessibilidade, as quais oferecem significativa contribuição ao processo de inclusão digital de pessoas com necessidades especiais, em particular cegos e surdos-mudos.

Finalmente, o trabalho de disseminação das idéias e técnicas desenvolvidas nesta pesquisa a outros pares de línguas assim como em artigos de periódicos e *corpora* temáticos também oferece um campo fértil de pesquisa e de aplicações práticas em favor da diversidade linguística no campo da ciência.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, Fábio. Esforço Cognitivo e Efeito Contextual em Tradução: Relevância no Desempenho de Tradutores Novatos e Expertos. **Linguagem em (dis)curso - Lemd**, Tubarão, v. 5, n. , p.11-31, 2005. Disponível em: <<http://linguagem.unisul.br/paginas/ensino/pos/linguagem/linguagem-em-discurso/0503/050301.pdf>>. Acesso em: 12 out. 2017
- ALVES, Fabio; GONÇALVES, José Luiz V. R.. A Relevance Theory approach to the investigation of inferential processes in translation. In: ALVES, Fabio (Org.). **Triangulating Translation**. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2003. Cap. 1. p. 3-24. Disponível em: <<https://pdfs.semanticscholar.org/a858/59ca739059c8585c2cd06eb53a12593a2b76.pdf?page=14>>. Acesso em: 12 out. 2017.
- ARANGO, J. Architectures. In: Journal of Information Architecture. Vol. 3, No. 1, p. 41-47, 2011. Semestral. Disponível em <http://journalofia.org/volume3/issue1/04-arango/> Acesso em: 26 out. 2015.
- BAKHTIN, Mikhail Mikhailovich. **The Dialogic Imagination: Four Essays** by M. M. Bakhtin. Austin: University Of Texas Press, 1981. (University of Texas Press Slavic series, no. 1). Disponível em: <http://www.public.iastate.edu/~carlos/607/readings/bakhtin.pdf>. Acesso em: 15 jan. 2013.
- BALAGE FILHO, Pedro Paulo; PARDO, Thiago A. S. Pardo; NUNES, M. G. Volpe. **Sumarização Autoática de Textos Científicos: Estudo de Caso com o Sistema GistSumm**. Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional - NILC - ICMC-USP, São Carlos, SP, Brasil. Agosto 2017. Disponível em <http://conteudo.icmc.usp.br/pessoas/taspardo/NILCTR0711-BalageEtAl.pdf>, Acesso em 16.10.2017
- BALLARD, William L; HALLIDAY, M. AK. **Language as social semiotic: The social interpretation of language and meaning**. London: Edward Arnold, 1978.
- BARTHES, Roland. **Elementos de Semiologia**. São Paulo: Editora Cultrix, 1964.
- BELL, Daniel. **The Coming of Post-Industrial Society: A Venture in Social Forecasting**. Harmondsworth: Penguin, 1976..
- BENTIVOGLI, Luisa *et al.* Neural versus phrase-based machine translation quality: a case study, **arXiv:1608.04631**, 2016.
- BOJAR, Ondrej *et al.* Findings of the 2016 conference on machine translation (wmt16). **Proceedings of the First Conference on Machine Translation**, v. 2: Shared Task Papers, p. 131–198, Berlin, Germany, aug. 11-12, 2016.

BRANCO, António *et al.* **The Portuguese Language in the Digital Age**. Berlin: Springer, 2012.

BROUDOUX, Evelyne; CHARTRON, Ghislaine; CHAUDIRON, Stéphane. L'architecture de l'information: quelle réalité conceptuelle ? **Études de Communication**, Lille, v. 41, n. 2013, p.13-30, jun. 2013. Disponível em: <<http://edc.revues.org/5379>>. Acesso em: 14 ago. 2017.

CARNEIRO, M. D. C. (2014). Sistema BRACorpus. Certificado de registro de programa de computador. Processo: 14181-0. Titularidade: Fundação Universidade de Brasília. Instituto Nacional da Propriedade Industrial – INPI. Republica Federativa do Brasil. Ministério do Desenvolvimento, Indústria e Comércio Exterior. Disponível em: <<http://let.unb.br/bracorpus/index.php/exemplo>>. Acesso em: 17 nov. 2016.

CASTELLS, Manuel. **A Sociedade em rede**. 11. ed. São Paulo: Paz e Terra, 2008.

CATIVELLI, Adriana Stefani. **Biblioteca Universitária**: serviços transformados pelas tecnologias. 2009. 53 f. Trabalho de Conclusão de Curso (Graduação em biblioteconomia) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Florianópolis, 2009. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/120213/284432.pdf?sequencia=1>>. Acesso em: 9 nov. 2015.

CHOMSKY, Noam. **Aspects of the Theory of Syntax**. Cambridge: MIT Press, 2014.

COOPER, Liane. **What's an Information Architect**. 2009. Disponível em: <<http://slistechnologyreview.pbworks.com/f/557+Cooper+Term+Paper+102909.docx>>. Acesso em: 26 out. 2015.

CRESWELL, John W. **Projeto de pesquisa**. Porto Alegre: Artmed, 2010.

DRUCKER, Peter. The next society. **The economist**, v. 52, nov. 1st 2001.

ElBaradei, Mohamed, **New York Times**, 10 de Fevereiro de 2011) *apud* Jorge Arango, in Journal of Information Architecture, vol. 3, Tomo 1, in Architectures, 2011

FODOR, Jerry A. **The modularity of mind**. Cambridge. Massachusetts: Mit Press, 1986.

FOX, Christopher. A stop list for general text, **ACM SIGIR Forum**. v. 24, p. 19-21, 1989.

FRANCA, A. I.; MAIA, Marcus (Org.). **Papers in Psycholinguistics**: Proceedings of the First International Psycholinguistics Congress, Anpoll's Psycholinguistics Work Group. Rio de Janeiro: Imprinta, 2010. 457 p.

FRIAS-MARTINEZ, E. et al. Automated user modeling for personalized digital libraries. **International Journal of Information Management**, v. 26, n. 3, p. 234-248, 2006.

GALE, William A.; CHURCH, Kenneth W. A program for aligning sentences in bilingual corpora. **Computational linguistics**, v. 19, n. 1, p. 75-102, 1993.

GAUCH, Susan et al. User profiles for personalized information access. In: BRUSILOVSKY, P.; KOBASA, A.; NEJDL, W. (Eds.): **The Adaptive Web**, LNCS 4321, 2007. p. 54-89. Disponível em <<http://www.dcs.warwick.ac.uk/~acristea/courses/CS411/2008/Book%20-%20The%20Adaptive%20Web/UserProfilesforPersonalizedInfoAccess.pdf>>. Acesso em: 15 jun. 2017.

GOLEMATTI, Maria et al. Creating an ontology for the user profile: Method and applications. In: **Proceedings of the first RCIS conference**. Ouarzazate, Marrocos. p. 407-412, apr. 23-26, 2007. Disponível em: <<http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Creating an Ontology for the User Profile Met>>. Acesso em: 26 set. 2015.

GRICE, H. P. **Studies in the way of words**. Cambridge: Cambridge University Press, 1989.

GUPTA, Vishal; LEHAL, Gurpreet Singh. A survey of text summarization extractive techniques. **Journal of emerging technologies in web intelligence**, v. 2, n. 3, p. 258-268, 2010.

GUTT, Ernst-august. Translation and Relevance. 1989. 309 f. Tese (Doutorado em Linguistic) - Departament Of Phonetis And Lingistics, University College London, London, 1989. Disponível em: <<http://discovery.ucl.ac.uk/1317504/1/241978.pdf>>. Acesso em: 15 nov. 2015.

HALLIDAY, M. A. K. **An Introduction to Functional Grammar**. 3. ed. London: Arnold, 2004.

HARDT, Michael; NEGRI, Antonio. **Empire**. Paris: Exils Éditeur, 2000.

HARDT, Michael et al. The global coliseum: on Empire. **Cultural Studies**, v. 16, n. 2, p. 177-192, 2002.

HILLMAN, Donald J. The notion of relevance (I). **Journal of the Association for Information Science and Technology**, v. 15, n. 1, p. 26-34, 1964.

HINTON, Andrew (Org.). The Machineries of Context. **Journal of Information Architecture**, Copenhagen, v. 1, n. 1, p. 37-47, 01 abr. 2009. Semestral. Disponível em: <<http://journalofia.org/volume1/issue1/04-hinton/jofia-0101-04-hinton.pdf>>. Acesso em: 26 out. 2015.

IEDEMA, Rick. Multimodality, resemiotization: extending the analysis of discourse as multi-semiotic practice. **Visual Communication**, London, v. 2, n. 1, p.29-57, fev. 2003

JEWITT, Carey. An Introduction to Multimodality. In: JEWITT, Carey. **The Routledge Handbook of Multimodal Analysis**. New York: Routledge, 2009. Cap. 1. p. 14-27.

JONES, Rodney H.. Technology and sites of display. In: JEWITT, Carey. **The Routledge Handbook of Multimodal Analysis**. New York: Routledge, 2009. Cap. 8. p. 114-126.

KRESS, Gunther. What is Mode ? In: JEWITT, Carey. **The Routledge Handbook of Multimodal Analysis**. New York: Routledge, 2009. Cap. 4. p. 54-67.

KRESS, Gunter; van LEEUWEN, Theo. **Multimodal Discourse: The modes and media of contemporary communication**. New York: Oxford University Press, 2001.

KRESS, Gunther; van LEEUWEN, Theo. **Multimodal Discourse**. Londres: Bloomsbury Academic, 2001. 142 p.

KRESS, Gunther; van LEEUWEN, Theo. **Reading Images: The Grammar of Visual Design**. 2. ed. New York: Routledge, 2006. 296 p.

LAZZARIN, Fabiana Aparecida et al. Da informação à compreensão: reflexões sobre Arquitetura da Informação, Usabilidade e Acessibilidade no campo da Ciência da Informação. **Biblioline**, João Pessoa, v. 8, n. , p.231-244, jan. 2012.

LEE-JAHNKE, Hanelore; DELISLE, Jean; CORNIER, Monique C. (organizadores); tradução e adaptação para o português [de] Alvaro Faleiros e Cláudia Xatara – Brasília: Editora Universidade de Brasília, 2013, 172p, ISBN 978-85-230-1071-3

LEVACOV, Marília. Tornando a informação disponível: o acesso expandido e a reinvenção da biblioteca. In: MARCONDES, Carlos H. et al (Org.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Brasília: Ibict, 2006. p. 205-221.

LLORET, Elena *et al.* Compendium: a text summarisation tool for generating summaries of multiple purposes, domains, and genres. **Natural Language Engineering**, v. 19, n. 2, p. 147-186, 2013.

LONGCHAMPS, Jacqueline Rodrigues. **Deficit Específico de Linguagem (DEL-PRAG) e dificuldades de aprendizagem: interface gramática-pragmática e relevância no uso da língua**. 2014, 244 f.: il; 30 cm. Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Letras, 2014.

LUHN, Hans Peter. The automatic creation of literature abstracts. **IBM Journal of research and development**, v. 2, n. 2, p. 159-165, 1958.

MARCONDES, Carlos H. et al (Org.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: Ufba, 2006.

MÁRDERO, Arellano. Àngel. Serviços de referência virtual. **Ciencia da Informação**, Brasília, v. 30, p.1-15, 2001.

MENEZES, Cláudio; BAPTISTA, Dulce Maria. Metodologia de Acesso a Dissertações de Mestrado de Tradução por Estrangeiros: Uma abordagem preliminar. **Revista Iberoamericana de Ciência da Informação**, Brasília, v.10, n.1, p. 154-163, jan./jul. 2017. Disponível em

<http://periodicos.unb.br/index.php/RICI/article/view/16462/18074>. Acesso em 16.10.2017

MENEZES, Francisco Cláudio Sampaio de. O Multilinguismo e as Novas Tecnologias das Línguas no Século XXI. **Belas Infiéis**, Brasília, v. 4, n. 12015, p.85-98, 01 jun. 2015. Disponível em: <<http://periodicos.unb.br/index.php/belasinfiéis/issue/view/1175/showToc>>. Acesso em: 15 nov. 2015.

MIHALCEA, R.; TARAU, P. TextRank: Bringing order into texts. Association for Computational Linguistics. **EECS News**, jul. 2004. Disponível em: <https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>. Acesso em: 19 jun. 2017.

Nationall Science Foundation, Washington, D.C. OT-SP-77 12(2)

NORRIS, Sigrid. **Analyzing Multimodal Interaction: A Methodological Framework**. London: Routledge, 2004.

O'HALLORAN, Kay. **Multimodal discourse analysis: Systemic functional perspectives**. London, Uk: A & C Black Publishers, 2004.

O'TOOLE, M. Opera Ludents: the Sidney Opera House at work and play. In: O'HALLORAN, Kay L., **Multimodal Discourse Analysis**. London: Continuum, 2004. p. 11-27.

PAYEUR, Cécile, Évaluation et conception d'un dispositif numérique de médiation documentaire sur support mobile dans l'espace museal. In: DISPOSITIFS NUMÉRIQUES: Contenus, Interactivité et Visualisation, CIDE 16, **Actes du 16ème Colloque Internatinal sur le Document Electronique**, Lille, 21-22 de novembro de 2013, ISBN 979-10-90094-14-7, Europa Productions, Paris, p. 115-123, nov. 2013.

PESSOA, Patrícia; CUNHA, Murilo Bastos da. Perspectivas dos serviços de referência digital. **Informação & Sociedade**, João Pessoa, v. 17, n. 3, p. 69-82, set. 2007. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000004782/944c9aa5caf9dd83f20e62713d0badbe>>. Acesso em: 15 jun. 2017.

PICKARD, Alison Jane. **Research methods in information**. Londres: Facet Publishing, 2007. 336 p

PORAT, Marc U. **The Information Economy; Sources and Methods for Measuring the Primary Information Sector (Detailed Industry Reports)**, 1977, Office of Telecommunications (DOC), Washington. Disponível em <http://files.eric.ed.gov/fulltext/ED142206.pdf>. Acesso em 16.10.2017

PORAT, Marc. U., **The Information Economy: Definition and Measurement**. 1977, Office of Telecommunications (DOC), Washington, D.C. National Science Foundation, Washington, D.C. OT-SP-77-12(1) <http://files.eric.ed.gov/fulltext/ED142205.pdf>, Acesso em 15.out.2017

RIBEIRO, Larissa O. **Sumarização Automática**: uma ferramenta para o tratamento automático das línguas naturais. 2016. 43 f., il. Trabalho de conclusão de curso (Bacharelado em Línguas Estrangeiras Aplicadas)—Universidade de Brasília, Brasília, 2016.

RINO, Lúcia Helena Machado et al. Summarizers of Texts in Brazilian Portuguese: Lecture Notes on Artificial Intelligence. In: 17TH BRAZILIAN SYMPOSIUM ON ARTIFICIAL INTELLIGENCE, 1., 2004, São Luis. **Proceedings of the 17th Brazilian Symposium on Artificial Intelligence**. São Luís: Springer-verlag, 2004. v. 1, p. 235 - 244. Disponível em: <https://www.researchgate.net/publication/220974768_A_Comparison_of_Automatic_Summarizers_of_Texts_in_Brazilian_Portuguese>. Acesso em: 29 set. 2004.

SANTOS, Liliane Simões dos. **Análise dos softwares Gistsummarizer e Google Tradutor**: ferramentas de suporte à sumarização de textos científicos e à tradução automática. 2014. 47 f. Trabalho de conclusão de curso (Bacharelado em Línguas Estrangeiras Aplicadas - LEA/MSI) – Universidade de Brasília. Instituto de Letras. 2014. Disponível em: <http://bdm.unb.br/bitstream/10483/9576/1/2014_LilianeSimoedosSantos.pdf>. Acesso em: 16 nov. 2015.

SAUSSURE, Ferdinand de. **Curso de Lingüística Geral**. São Paulo: Cultrix, 1975.

SARACEVIC, Tefko. Relevance: A review of and a framework for the thinking on the notion in information science. **Journal of the American Society for information science**, v. 26, n. 6, p. 321-343, 1975.

SARACEVIC, Tefko. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. **Journal of the American Society for information Science and Technology**, v. 58, n. 13, p. 2126-2144, 2007.

SCHAMBER, Linda; EISENBERG, Michael. Relevance: The Search for a Definition. In: 51st Annual Meeting of the American Society for Information Science, 51. 1988, Atlanta. **Proceedings...** . Atlanta: Educational Resources Information Center, 1988. v. 1, p. 1 - 17. Disponível em: <<http://eric.ed.gov/?id=ED304158>>. Acesso em: 01 nov. 2015.

SCOLLON, Ron; SCOLLON, Suzy Wong. **Discourses in Place**: Language in the Material World. New York: Routledge, 2003.

SEARLE, John. **Speech Acts**: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press, 1969.

SHANNON, Claude; WEAVER, Warren. **The Mathematical Theory of Communication**. Verlag: University Of Illinois Press, 1949.

SHANNON, Claude; WEAVER, Warren. **The mathematical theory of communication**. Urbana: University Of Illinois Press, 2015.

SILVA, Patrick Pereira. **ExtraWeb**: um sumarizador de documentos Web baseados em etiquetas HTML e ontologia, 2006. 168 f. Dissertação (Mestrado em Ciências Exatas e da Terra) - Universidade Federal de São Carlos, São Carlos, 2006.

SOUZA, C.F.R.; NUNES, M.G.V. **Avaliação de Algoritmos de Sumarização Extrativa de Textos em Português**. Relatórios Técnicos do ICMC-USP. NILC-TR-01-09, Novembro 2001

SPERBER, Dan; WILSON, Deirdre. Introduction: pragmatics. In: WILSON, Deirdre; SPERBER, Dan. **Meaning and Relevance**. Cambridge: Cambridge University Press, 2002. p. 1-27, ISBN 978-0-521-74748-6

SPERBER, Dan; WILSON, Deirdre. Pragmatics, Modularity and Mind-Reading. **Mind And Language**, London, n. 17, p.3-23, 2012. Disponível em: <<http://www.dan.sperber.fr/wp-content/uploads/2009/09/PragmaticsModularityMindReading.pdf>>. Acesso em: 7 nov. 2015.

SPERBER, Dan; WILSON, Deirdre. **Relevance**: communication and cognition. 2. ed. Oxford: Blackwell, 1995 pp vii + 326

THONG, James YL; HONG, Weiyin; TAM, Kar Yan. What leads to user acceptance of digital libraries?. **Communications of the ACM**, v. 47, n. 11, p. 78-83, 2004.

TRAJKOVA, Joana; GAUCH, Susan. Improving Ontology-Based User Profiles. **Proceeding Riao 2004**, Avignon, v. 1, n. 1, p.380-390, 30 abr. 2004. Anual. Disponível em: <<http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/ImprovingOntology-BasedUserProfiles.pdf>>. Acesso em: 20 out. 2015.

UNESCO, “Recommandation sur la Promotion et l’Usage du Multilinguisme et l’Accès Universel au Cyberspace”, Disponível em: <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/official_documents/Fre%20-%20Recommendation%20concerning%20the%20Promotion%20and%20Use%20of%20Multilingualism%20and%20Universal%20Access%20to%20Cyberspace.pdf>, 2003, Acesso em: 05 out..2017

UNESCO, “A Decade in Promoting Multilingualism in Cyberspace”, Disponível em <http://unesdoc.unesco.org/images/0023/002327/232743e.pdf>, Acesso em: 05 out. 2017

VAN LEEUWEN, Theo. 21 Multimodality. *The Handbook of Discourse Analysis*, 2015, p. 447. Disponível em http://www.philsci.univ.kiev.ua/UKR/courses/asp/asp-lit/tannen_d_hamilton_h_e_schiffrin_d_eds_the_handbook_of_discou.pdf#page=477, Acesso em 17.08.2017

WEBSTER, Franck. **Theories of the Information Society**. 3. ed. London: Routledge, 2006.

WILSON, Deirdre; SPERBER, Dan. Teoria da Relevância. **Linguagem em (dis)curso**, Tubarão, v. 5, n. especial, p. 221-268, 2005. Disponível em http://portaldeperiodicos.unisul.br/index.php/Linguagem_Discurso/article/viewFile/287/301. Acesso em 16.10.2017

WILSON, Deirdre; SPERBER, Dan (Org.). Pragmatics and Cognition. In: HORN, Lawrence R.; WARD, Gregory. **The Handbook of Pragmatics: Relevance Theory**. 2. ed. Malden: Blackwell Publishing, 2004. Cap. 4. p. 607-632.

WILSON, Deirdre; SPERBER, Dan. **Relevance Theory**. 2002. Disponível em: <http://cogprints.org/2317/1/relevance_theory.htm#_edn1>. Acesso em: 02 nov. 2015.

WILSON, Tom D. On user studies and information needs. *Journal of documentation*, 1981, vol. 37, no 1, p. 3-15.

WU, Yonghui et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.

ANEXO A – PRÉ-TESTE – QUESTIONNAIRE 1

Note d'introduction

Ce questionnaire a pour objectif de collecter des informations sur les obstacles rencontrés par les étudiants étrangers en ce qui concerne la localisation et la compréhension des textes dans des sites web écrits en Portugais. Le questionnaire est structuré en 4 points:

- 1) Identification
- 2) Questions générales sur l'accès par les interviewés aux sites web dans leur langue maternelle et en Portugais;
- 3) Questions spécifiques sur le site web de l'UnB;
- 4) Questions techniques sur les modes (interfaces) utilisés dans les sites web par les répondants.

1 - Identification

Nom, Prénom

Numéro d'inscription à l'UnB

Cours à l'UnB:

Nationalité:

Age:

Niveau d'études:

Date d'arrivée au Brésil:

2 – Questions générales

1. Vous accédez à des sites web dans votre langue maternelle ?

1. OUI
2. NON

2. Quand vous accédez à ces sites web, obtenez vous facilement l'information demandée dans votre langue maternelle ?

1. OUI
2. NON

3. Pour quelles raisons ?

4. Rencontrez-vous des difficultés pour localiser l'information recherchée dans des sites web en Portugais du Brésil ?

1. OUI
2. NON

5. Si oui, lesquelles ?

6. Avez-vous des difficultés pour comprendre l'information obtenue dans des sites web en Portugais du Brésil ?

7. Si oui, le(s)quel(les) ?

3 – Questions sur le site web de l'UnB
--

1. Avez vous déjà accédé au site web de l'UnB ?

1. OUI
2. NON

2. Quelles sections du site web de l'UnB connaissez-vous ?

3. Quelles sections du site web de l'UnB (citées ci-dessus) avez-vous visité ?

4. Connaissez-vous le Répertoire Institutionnel de l'UnB ?

1. OUI
2. NON

5. Avez-vous déjà accédé à des ressources du Répertoire Institutionnel ?

1. OUI
2. NON

6. Avez-vous déjà consulté un mémoire de fin de cursus ?

1. OUI
2. NON

7. Avez-vous déjà consulté un mémoire de master ?

1. OUI
2. NON

8. Você já consultou alguma Tese de Doutorado ? Avez vous déjà consulté une thèse de doctorat ?

1. OUI
2. NON

4 –Questions techniques

1. Cochez les options (modes) que vous avez rencontré en utilisant des sites web dans votre langue maternelle:

1. Vidéo
2. Tutorial de voix
3. Tutorial descriptif écrit
4. Images d'illustration du contenu.
5. Autres

2. Cochez les cases avec l'option(s) qui vous avez rencontré en utilisant des sites web écrit en Portugais du Brésil:

1. Video
2. Tutorial de voix
3. Tutorial descriptif écrit
4. Images illustratives du contenu
5. Autres

3. Avez-vous déjà utilisé des logiciels informatiques de traduction automatique pour traduire des contenus de sites web brésiliens à votre langue maternelle ?

1. OUI
2. NON

4. Si oui, le(s)quel(s) ?

5. Avez-vous déjà utilisé des logiciels informatiques pour l'élaboration automatique de sommaires de textes en portugais brésilien ?

1. OUI
2. NON

6. Si oui, lequel(s) ?

7. Avez-vous déjà utilisé une quelconque ressource informatique (logiciels informatiques, applications, etc) pour vous aider dans la compréhension des contenus écrits en portugais brésilien ?

1. SIM
2. NON

8. Si oui, lequel(s)

9. Pouvez vous nous donner des précisions sur l'usage de technologies d'information et communications dans votre expérience en essayant de comprendre des contenus numériques en Portugais du Brésil ?

10. Avez vous eu des difficultés pour comprendre les questions de ce questionnaire ?

11. Si oui, le(s)quel(les) ?

ANEXO B – PRÉ-TESTE

QUESTIONÁRIO QUESTIONNAIRE

Introdução Note d'introduction

Este questionário tem por objetivo coletar informações com estudantes estrangeiros sobre as dificuldades encontradas na localização de informação e na compreensão de seu significado em websites escritos em Português. O questionário está estruturado em 4 seções:

- 1) Identificação do entrevistado
- 2) Questões gerais sobre o acesso dos entrevistados a sites no seu país de origem e no Brasil;
- 3) Questões específicas sobre o site da UnB;
- 4) Questões técnicas sobre os modos (interfaces) usados em sites web usados pelos entrevistados.

Ce questionnaire a pour objectif de collecter des informations sur les obstacles rencontrés par les étudiants étrangers en ce qui concerne la localisation et la compréhension des textes dans des sites web écrits en Portugais. Le questionnaire est structuré en 4 points:

- 1) Identification
- 2) Questions générales sur l'accès par les interviewvés aux sites web dans leur langue maternelle et en Portugais;
- 3) Questions spécifiques sur le site web de l'UnB;
- 4) Questions techniques sur les modes (interfaces) utilisés dans les sites web par les répondants.

1 – Identificação do Entrevistado 1 - Identification

Nome completo:
Nom, Prénom

Inscrição na UnB:
Numéro d'inscription à l'UnB

Curso na UnB:
Cours à l'UnB:

Nacionalidade:
Nationalité:

Idade:

Age:

Escolaridade:

Niveau d'études:

Data de chegada ao Brasil:

Date d'arrivée au Brésil:

2 - Questões gerais, 2 – Questions générales
--

1. Você acessa websites na sua língua materna ?

1. Vous accédez à des sites web dans votre langue maternelle ?

1. SIM, OUI
2. NÃO, NON

2. Ao acessar esses sites, você tem facilidade de obter a informação procurada na sua língua materna ? Quand vous accédez à ces sites web, obtenez vous facilement l'information demandée dans votre langue maternelle ?

1. SIM, OUI
2. NÃO, NON

3. Por quais razões ? Pour quelles raisons ?

4. Você encontra dificuldades para localizar a informação procurada em websites em português brasileiro ? Rencontrez-vous des difficultés pour localiser l'information recherchée dans des sites web en Portugais du Brésil ?

1. SIM, OUI
2. NÃO, NON

5. Quais ? Lesquelles ?

6. Você enfrenta dificuldades em compreender a informação obtida em websites em português brasileiro ? Avez-vous des difficultés pour comprendre l'information obtenue dans des sites web en Portugais du Brésil ?

7. Quais ? Lesquelles ?

3 – Questões sobre o website da UnB, Questions sur le site web de l'UnB

1. Você já acessou o website da UnB ? Avez vous déjà accédé au site web de l'UnB ?

1. SIM, OUI
2. NÃO, NON

2. Quais as seções do site da UnB que você conhece ? Quelles sections du site web de l'UnB connaissez-vous ?

3. Quais seções do site da UnB (acima citadas) você visitou ? Quelles sections du site web de l'UnB (citées ci-dessus) avez-vous visité ?

4. Você conhece o Repositório Institucional da UnB ? Connaissez-vous le Répertoire Institutionnel de l'UnB ?

1. SIM
2. NÃO

5. Você já acessou algum recurso do Repositório Institucional ? Avez-vous déjà accédé à des ressources du Répertoire Institutionnel ?

1. SIM, OUI
2. NÃO, NON

6. Você já consultou alguma Monografia de Graduação ? Avez-vous déjà consulté un mémoire de fin de cursus ?

7. Você já consultou alguma Dissertação de mestrado ? Avez-vous déjà consulté un mémoire de master ?

1. SIM, OUI
2. NÃO, NON

8. Você já consultou alguma Tese de Doutorado ? Avez vous déjà consulté une thèse de doctorat ?

1. SIM, OUI
2. NÃO, NON

4 – Questões técnicas, Questions techniques

1. Marque as opções (modo) que você encontrou ao usar websites na sua língua maternal: Cochez les options (modes) que vous avez rencontré en utilisant des sites web dans votre langue maternelle:

1. Video, video
2. Tutorial de voz, tutorial de voix
3. Tutorial descritivo escrito, tutorial descriptif écrit
4. Imagens ilustrativas do conteúdo, images d'illustration du contenu.
5. Outros, Autres

2. Marque as opções que você encontrou ao usar websites escritos em Português do Brasil: Cochez les cases avec l'option(s) qui vous avez rencontré en utilisant des sites web écrit en Portugais du Brésil:

1. Video, Video
2. Tutorial de voz, Tutorial de voix
3. Tutorial descritivo escrito, Tutorial descriptif écrit
4. Imagens ilustrativas do conteúdo, Images illustratives du contenu
5. Outros, Autres

3. Você já usou algum programa de computador para **tradução automática** para traduzir conteúdos de websites brasileiros para a sua língua materna? Avez-vous déjà utilisé des logiciels informatiques de traduction automatique pour traduire des contenus de sites web brésiliens à votre langue maternelle ?

1. SIM, OUI
2. NÃO, NON

4. Se sim, quais ? Si oui, le(s)quel(s)?

5. Você já usou algum programa de computador para **sumarização automática** de textos em português brasileiro ? Avez-vous déjà utilisé des logiciels informatiques pour l'élaboration automatique de sommaires de textes en portugais brésilien ?

1. SIM, OUI
2. NÃO, NON

6. Se sim, quais ? Si oui, lequel(s) ?

7. Você já usou algum recurso de informática (programa de computador, aplicativo, etc.) para ajuda-lo na compreensão de conteúdos escritos em português brasileiro ? Avez-vous déjà utilisé une quelconque ressource informatique (logiciels informatiques, applications, etc) pour vous aider dans la compréhension des contenus écrits en portugais brésilien ?

1. OUI, SIM
2. NÃO, NON

8. Se sim, quais ? Si oui, lequel(s) ?

9. Você poderia listar alguns elementos sobre o uso das tecnologias de informação e comunicação na sua tentativa de compreender conteúdos digitais no Português do Brasil ? Pouvez vous nous donner des précisions sur l'usage de technologies d'information et communications dans votre expérience en essayant de comprendre des contenus numériques en Portugais du Brésil ?

10. Você teve dificuldade de compreender as questões deste questionário ? Avez vous eu des difficultés pour comprendre les questions de ce questionnaire ?

11. Se sim, quais ? Si oui, le(s)quel(les) ?

ANEXO C – PRÉ-TESTE – TABULAÇÃO DO QUESTIONÁRIO 1 (TABULATION DU QUESTIONNAIRE 1)

Cours à l'UnB	Nationalité	Age	Scolarité	Date d'arrivée au Brésil
Genie de Production	Congolaise, RDC	20		27/02/15
Ingénierie Electrique	Congolais	26	BAC	27/01/15
Langue Portugaise pour Étranger	Beninoise	21	Deux Annés Universitaire	11/03/15
NEPPE	Congolaise	23	Gradué	27/01/15
Lettres-Portugais--Anglais	Gabonais	22	Baccalauréat	02/02/15
NEPPE	Senegalais	24	BAC	25/02/15
NEPPE	Beninoise	26	Apprentissage de la langue portugaise	04/03/15
Português para Estrangeiros	Congolaise (R.A.C)	20	Diplomé d'Etat	27/01/15
Lettres-Traduction	Beninoise	26	Etudes Superieur	04/03/15
Langue Portugaise pour les Étrangers	Haïtienne	21	Universitaires	26/02/15
Português para Estrangeiros	Senegalais	27	Universitaire	01/05/15

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)

(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
1	O	O	Pour la recherche de connaissance basique de la lanngue portugaise	O	Le plus souvent, l'information rechercheé ne donne pas un résultat adéquat	Pas tellement, elle est abordable	N
2	O	N	Parce que le site me donne plusieurs réponses ou solution a ma demande alors je dois encore choisir la meilleur réponse parmi tant d'autres	O	La langue déjà est une difficulté; Plusieurs resultats sont proposés alors qu'il faut trouver le meilleur	Oui	Les anotations sur le panaux publique par exemple; Interpreter une carte pour localiser avec precision un endroid

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
3	O	O	La langue française que est ma langue maternel est l'une des langues les plus parlée de par le monde donc c'est normal que nous rencontrons sur la net une diversité	O	A vrai dire, nous parlons de deux langues et la construction des deux langues sont differente. Cela est un facteur que peut provoquer des difficultés dans la localisation de l'information	Peut arriver qu'il est un problème de comprehension mais le plus souvent on arrive a tirer l'important	Oui, dans le cas ou l'information presente des termes que sont un peu difficile a comprendre

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
4	O	O	Parfois dans la langue maternelle je comprends facilement et cela me permet a avoir plus d'expressions gramatical.	O	Quant nous interrogeons la machine parfois on a des bons resultats mais de fois le système donne autre chose que l'on a plus besoin	Pas de réponse	Les informations peuvent être compliqueer et l'utilisation aura encore un deuxième travail de consulter le dictionnaire pour comprendre le sens de mots compliquer.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
5	O	O	Ma langue maternelle, le français, est une langue internationale, langue de travail de l'ONU, ce qui fait que toutes les sciences connues de l'homme sont rédigées en français.	N	Pas de réponse	Le portugais est une langue latine comme le français. Il est vrai que le portugais du Brésil a une certaine particularité, il en va de même pour le français canadien. Je peux ne pas comprendre certains termes.	Pas de réponse
6	N	N	Des fois c'est difficile de trouver l'information chercher !	O	Bon, personnellement je pense que c'est des problèmes de compréhension de la langue portugaise.	Pas de réponse	A mon avis l'utilisation des mots et des phrases sont très élevées.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
7	O	O	Pas de réponse	O		Oui	Complexités des sites web; Problème de compréhension
8	O	O	Parce que je n'ai pas des difficultés à comprendre ma langue maternelle (Français)	N	Pour des raisons de compréhension de la langue,	Oui, quelques fois	Des fois, il m'est difficile de comprendre clairement et exactement ce qui est écrit généralement à cause des mots inconnus.
9	O	O	J'obtiens les informations facilement parce que j'ai toujours parlé français et donc j'ai une facilité de compréhension.	O	Pas de réponse	Oui, j'ai des difficultés pour comprendre l'information obtenue; parce que je ne comprends pas encore très bien le portugais.	J'ai des difficultés grammaticales et orthographiques.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

2. QUESTIONS GENERALES							
N. Question	Accès Web	Facilité dans la Langue Maternelle	Raisons pour l'obtention de l'Information en Français ?	Info en Portugais - Difficultés	Liste de Difficultés	Difficultés de compréhension	Liste de Difficultés
10	O	O	Pas de réponse	N	Je rencontre surtout de difficulté dans la compréhension ; parce que je ne parle pas encore très bien le portugais.	Non	Pas de réponse
11	O	O	Pour mieux connaître ma langue maternelle et pour mieux parler cette langue	Pas de réponse	Pas de réponse	Pas de réponse	

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

3. QUESTIONS SUR LE SITE WEB DE L'UnB								
N. Question	Accès Web	UnB: Sections du site web connues	UnB: Sections du site web visitées	UnB: Répertoire Institutionnel (RI)	Accès au RI	Consultation aux Mémoires de fin de cursus	Consultation aux Mémoires de Master	Consultation aux Thèses de Doctorat
1	O	Les sections qui concernent les pec-g ou la reconnaissance de NEPPE	Pas de réponse	N	N	N	N	N
2	O	Restaurant de l'UnB; Centre Olympique	Centre Olympique; Restaurant de l'UnB	N	N	O	N	N
3	N	Pas de réponse	Pas de réponse	N	N	N	N	N
4	O	Juste le site de bibliothèque pour voir les ouvrages du grammaire	Pas de réponse	N	N	N	N	P.R.
5	N	Pas de réponse	Pas de réponse	N	N	N	N	N
6	N	Pas de réponse	Pas de réponse	N	N	N	N	N

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

3. QUESTIONS SUR LE SITE WEB DE L'UnB								
N. Question	Accès Web	UnB: Sections du site web connues	UnB: Sections du site web visitées	UnB: Répertoire Institutionnel (RI)	Accès au RI	Consultation aux Mémoires de fin de cursus	Consultation aux Mémoires de Master	Consultation aux Thèses de Doctorat
7	O	La page d'accueil	Section informativ	N	N	N	N	N
8	O	La page initiale, la section des programmes des cours, la section sur le RU	Toutes	N	N	N	N	N
9	O	Je connais la section de "Alunos de Graduação"	Section "Biblioteca", Section "Post- Graduação"	N	N	N	N	N
10	O	Section des lettres	Section Lettre	NR	NR	N	N	N
11	N	Pas de réponse	Pas de réponse	N	N	N	N	N

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 1)						
N. Question	Web en Français: Ressources (Modes)	Web en Portugais: Ressources (Modes)	Usage de Logiciels de traduction automatique	Liste de Logiciels	Usage de logiciels d'élaboration automatique de sommaires	Liste de Logiciels
1	Video	Video, Tutorial descriptif écrit	O	Dictionnaire portugês traducteur	O	Google traduction
2	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	O	Google tradutor	N	P. R.
3	Video	Video	O	Google tradução, Dicionário informal, Bing Tradutor	N	P. R.
4	Video	Video	O	Quand je suis arrivé au Brésil je ne connaissais rien du portugais pour communiquer parfois j'utilisais le google traducteur	O	P. R.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 1)						
N. Question	Web en Français: Ressources (Modes)	Web en Portugais: Ressources (Modes)	Usage de Logiciels de traduction automatique	Liste de Logiciels	Usage de logiciels d'élaboration automatique de sommaires	Liste de Logiciels
5	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	O	Speak it, Voz wiki, Google tradutor	N	P. R.
6	Video, Images d'illustration du contenu, Autres	Video, Images illustratives du contenu	N	P.R.	N	P. R.
7	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	Video, Tutorial de voix, Tutorial descriptif écrit	O	Google traduction, Yandex, Dictionnaire Français-Portugais	O	Google traduction, Yandex
8	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu, Autres	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu, Autres	O	Le traducteur du navigateur Google Chrome	O	Google Traducteur

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 1)						
N. Question	Web en Français: Ressources (Modes)	Web en Portugais: Ressources (Modes)	Usage de Logiciels de traduction automatique	Liste de Logiciels	Usage de logiciels d'élaboration automatique de sommaires	Liste de Logiciels
9	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu	O	Google Traductor	N	P. R.
10	Video, Tutorial de voix, Tutorial descriptif écrit, Images d'illustration du contenu, Autres	Video, Tutorial de voix	O	Google traducteur	N	P. R.
11	Video	P.R.	P.R.	P.R.	P.R.	P. R.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 2)					
N. Question	Usage de Logiciels d'aide à la compréhension	Liste de Logiciels	Usage des TICs pour la compréhension de contenus numériques	Difficultés pour comprendre les questions de ce questionnaire ?	Liste des difficultés
1	O	Dictionnaire portugês, Duoling, Vocabulary trainer	Personnellement parlant, j'ai beaucoup apprécié	Non, parce que c'était bien claire pour comprendre	P. R.
2	O	Google tradutor	Oui la technologie m'a beaucoup aidé dans la compréhension de la langue avec l'application: conjuga-me.lnet.br et sinonimo.net.br	Non	P. R.
3	O	Tradutor portugês FR-PT Dictionary	L'usage de technologies au debut de mon apprentissage a été util mais après tu dois t' éloigner de ce moyen que peut porter prejudice dans le future	Non	P. R.
4	O	J'ai déjà utilisé um site conjuga-me, Youtube	Parfois j'utilisais le youtube pour chercher de cours online	J'ose dire non toutes les questionnaires sont simples, claires et compréhensibles	P. R.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 2)					
N . Question	Usage de Logiciels d'aide à la compréhension	Liste de Logiciels	Usage des TICs pour la compréhension de contenus numériques	Difficultés pour comprendre les questions de ce questionnaire ?	Liste des difficultés
5	O	Google tradutor	Il existe certains logiciels ou programmes uniquement pour le Portugais du Brésil. La localisation GPS nous propose de configurer nos appareils selon le système brésilien. Plusieurs applications ou logiciels nous sont proposés automatiquement,	Non	P. R.
6	N	Google tradutor	Il existe certains logiciels ou programmes uniquement pour le Portugais du Brésil. La localisation GPS nous propose de configurer nos appareils selon le système brésilien. Plusieurs applications ou logiciels nous sont proposés automatiquement,	Non	P. R.

ANEXO C - PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 2)					
N. Question	Usage de Logiciels d'aide à la compréhension	Application: Dicionario Português-Francês, Google Traductor	Je peut parler de "Dicionario Português"	Non. Il n'ai eu aucune difficulté a comprendre des questions du questionnaire.	
7	O	P.R	Quand j'ai du mal a comprendre quelque chose em portugais j' esseye de faire la traduction dans ma langue ou a travers des applications téléphoniques	Não	P. R.
8	N	dictionaire	Pour une personne qui apprend une langue donnée l' usage de cette technologie est essentiel car cela aide beaucoup à apprendre	Pas du tout	P. R.
9	O	Application: Dicionario Português-Francês, Google Traductor	Je peut parler de "Dicionario Português"	Non. Il n'ai eu aucune difficulté a comprendre des questions du questionnaire.	
10	O	Wikiconario PT	J'utilize souvent la technologie pour améliorer mês connaissances dans une langue étrangère.	Non	P. R.

ANEXO C- PRÉ-TESTE - TABULAÇÃO DO QUESTIONÁRIO 1 (Tabulation du Questionnaire 1)
(cont)

4. QUESTIONS TECHNIQUES (page 2)					
N. Question	Usage de Logiciels d'aide à la compréhension	Application: Diccionario Portuguêês-Francês, Google Traductor	Je peut parler de "Diccionario Portuguêês"	Non. Il n'ai eu aucune difficulté a comprendre des questions du questionnaire.	
11	P.R.	P.R	J'utilize informatique pour chercher par exemple la signification des mots, e pour faire des recherches au certaines informations.	Souvent je rencontre des difficultés pour comprendre les questions ou parfois je ne trouve l' infomration cherché.	

ANEXO D – QUESTIONÁRIO 2



GROUPE D'ÉTUDES ET DE RECHERCHE INTERDISCIPLINAIRE EN INFORMATION ET COMMUNICATION

PPG | PROGRAMA DE PÓS GRADUAÇÃO EM
CINF | CIÊNCIA DA INFORMAÇÃO

 UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO

QUESTIONNAIRE

Formulation de Concepts pour la Création d'une Méthodologie d'accès à des Contenus en Portugais par des Étrangers

Note d'introduction

Ce questionnaire a pour objectif de collecter des informations sur les principaux obstacles rencontrés par les étudiants de langue maternelle française dans la recherche et la compréhension des textes dans des sites web écrits en Portugais, particulièrement des textes de littérature scientifique.

Parmi les nombreuses ressources numériques de littérature scientifique en langue portugaise on a choisi de faire cette recherche à partir des ressources numériques suivants:

- Biblioteca Digital de Teses e Dissertações (Bibliothèque numérique de thèses et dissertations, BDTD, <http://bdtd.ibict.br>, Brésil);
- Diretório Luso-Brasileiro (Répertoires de Accès libre OASIS (<http://oasisbr.ibict.br>, Brésil) et Répertoire scientifique d'Accès ouvert de Portugal, RCAPP (<https://www.rcaap.pt>, Portugal);
- Biblioteca Digital de Monografias de Graduação da Universidade de Brasília (Mémoires de fin de cursus, <http://bdm.unb.br>, UnB/RiUnB);
- Dissertações de Mestrado da Universidade de Brasília (Mémoires de Master, http://repositorio.unb.br/?locale=pt_BR, UnB/RiUnB);
- Teses de Doutorado da Universidade de Brasília (http://repositorio.unb.br/?locale=pt_BR, Thèses de Doctorat de l'Université de Brasília, UnB/RiUnB).
- SCIELO ("Scientific Electronic Library On-Line", <http://www.scielo.org>).

On demande à chaque étudiant de visiter 2 sites web:

- 1) Bibliothèque Numérique de Thèses et Dissertations (BDTD, <http://btdt.ibict.br>);
- 2) Un deuxième site web, choisi parmi les suivants:
 - a) SCIELO, <http://www.scielo.org>;
 - b) OASIS, <http://www.oasisbr.ibict.org>;
 - c) RCAAP, <http://www.rcaap.pt>;
 - d) Bibliothèque Numérique de Mémoires de fin de cursus de l'Université de Brasília, <http://bdm.unb.br>;
 - e) Bibliothèque Numérique de Dissertations de Master de l'Université de Brasília, http://repositorio.unb.br/?locale=pt_BR;
 - f) Bibliothèque Numérique de Thèses de Doctorat de l'Université de Brasília, http://repositorio.unb.br/?locale=pt_BR.

Dans le premier site web, la BDTD, on demande à l'étudiant de choisir une thèse ou une Mémoire de Master et d'essayer de comprendre son contenu. Pour le deuxième site web à visiter, il/elle devra choisir un document de son intérêt et essayer de comprendre son contenu.

Ce questionnaire est structuré en 4 parties:

- 1) Identification du répondant;
- 2) Questions générales concernant l'accès des interviewés aux sites web en français et en portugais;
- 3) Questions spécifiques relatives à des sites web de contenu scientifique en portugais
- 4) Questions techniques relatives aux modes (interfaces) des sites web utilisés par les interviewés.

1 – Identification

Nom, Prénom:

Institution:

Inscription dans l'Institution:

Cours:

Niveau du Cours (Bac, Master, Doctorat):

Nationalité:

Age:

Nombre d'années d'étude du Portugais:

2 – Questions générales

3. Parlez-vous portugais dans votre cercle familial ?

1. () OUI
2. () NON

4. Avez vous le diplôme CELPE-BRAS ou un diploma équivalent de proficiencia en Portugais

1. () OUI
2. () NON

5. Si oui, depuis combien de temps ?

1. () Moins d'un an
2. () Entre 5 et 10 ans
3. () Plus de 5 ans.

6. Accédez-vous à des sites web dans votre langue maternelle ?

1. () OUI
2. () NON

7. Si oui, quand vous accédez à ces sites web, obtenez-vous facilement l'information demandée dans votre langue maternelle ?

1. () OUI
2. () NON

5. Si oui, cocher quelles sont les raisons qui expliquent la facilité de cet accès

1. () Disponibilité d'un menu pour le choix de la langue d'usage;

- 2. () Existence d'une rubrique spécifique pour l'accès à des étrangers
- 3. () Page d'ouverture avec des illustrations adéquates à l'usage par des étrangers;
- 4. () Description orale dans votre langue maternelle;
- 5. () Vidéo détaillant les différentes rubriques du site web et de son contenu dans votre langue maternelle
- 6. () Autres
- 7. () Aucune raison spécifique.

6. Sur la base de la description des ressources numériques qui ont été présentés dans l'introduction de ce questionnaire (où il est demandé à chacun/une de visiter la BDTD et un deuxième site web), rencontrez-vous des difficultés pour accéder à l'information recherchée dans des sites web en Portugais ?
- 1. () OUI
 - 2. () NON

7. Si oui, décrivez brièvement quelles sont ces difficultés.

8. Une fois que vous avez accédé au document recherché, avez-vous des difficultés pour comprendre l'information obtenue dans des sites web en Portugais ?
- 1. () OUI
 - 2. () NON

9. Lesquelles ?

- 1. () Absence d'un menu pour le choix de la langue d'usage;
- 2. () Absence d'une rubrique spécifique pour l'accès pour des étrangers;
- 3. () Absence de figures illustratives adéquates à l'usage par des étrangers dans la page d'ouverture du site;
- 4. () Absence d'une description orale dans ma langue maternelle, avec des explications sur les fonctionnalités et contenus du site web;

5. () Absence d'une vidéo explicative des fonctionnalités du site web et de son contenu dans ma langue maternelle;

6. () Autres

7. () Aucune raison spécifique

3 – Questions sur les ressources numériques d'information scientifique en Portugais

1. Avez-vous consulté un mémoire de master dans la Bibliothèque Numérique de Thèses et Mémoires ?

1. () OUI

2. () NON

2. Si oui, avez-vous compris le contenu du mémoire consulté ?

1. () OUI

2. () NON

3. Avez-vous consulté une Thèse de Doctorat de la Bibliothèque Numérique de Thèses et Dissertations ?

1. () OUI

2. () NON

4. Si oui, avez vous compris le contenu de la Thèse consultée ?

1. () OUI

2. () NON

5. Citer le nom de la deuxième ressource numérique que vous avez consultée.

6. Connaissez vous le répertoire Institutionnel de la Université de Brasília (UnB) ?

1. () OUI

2. () NON

7. Si oui, avez vous eu accès à au moins un document sur la liste du répertoire Institutionnel de l'UnB ?

1. () OUI

2. () NON

8. Si oui, avez-vous compris le contenu de ce document ?
1. () OUI
 2. () NON
9. Avez-vous consulté des Mémoires de fin de cursus dans la base de données ou le répertoire Institutionnel de l'UnB ?
1. () OUI
 2. () NON
10. Si oui, avez-vous compris le contenu du document consulté ?
1. () OUI
 2. () NON
11. Avez-vous consulté un Mémoire de Master disponible sur la liste du répertoire Institutionnel de l'UnB ?
1. () OUI
 2. () NON
12. Si oui, avez-vous compris le contenu du Mémoire consulté ?
1. () OUI
 2. () NON
13. Avez-vous déjà consulté une Thèse de Doctorat disponible sur la liste du répertoire Institutionnel de l'UnB ?
1. () OUI
 2. () NON
14. Si oui, avez-vous compris le contenu de la Thèse consultée ?
1. () OUI
 2. () NON
15. SVP, Indiquer ci-après le titre des deux documents qui vous avez consulté.
- 1.
 - 2.

4 – Questions techniques

1. Cochez les ressources que vous avez rencontrées dans l'usage de sites web dans votre langue maternelle:
1. () Video
 2. () Description orale

3. () Texte explicatif
4. () Images illustratives du contenu
5. () Autres

2. Cocher les options que vous avez rencontrées dans l'usage de sites web écrits en Portugais.

1. () Video
2. () Description orale
3. () Texte explicatif
4. () Images illustratives du contenu
5. () Autres

3. Avez-vous déjà utilisé des logiciels de résumé automatique de textes en portugais ?

1. () OUI
2. () NON

4. Avez-vous déjà utilisé des logiciels pour la traduction automatique de la langue (TAL) pour traduire des contenus de sites web en portugais pour votre langue ?

1. () OUI
2. () NON

5. Avez-vous déjà utilisé une quelconque ressource informatique (logiciels, dictionnaire ou terminologie électronique, thésaurus informatisé, logiciel de fouille de textes, traduction automatique, autres) pour vous aider dans la compréhension de contenus écrits en portugais ?

1. () OUI
2. () NON

11. Si oui, avez vous compris le résultat produit par le logiciel de traduction automatique de la langue ?

1. () OUI
2. () NON

12. Indiquer sur une échelle de 1 à 5 l'utilité de ce logiciel de traduction pour la compréhension du document traduit.

1. ()
2. ()
3. ()
4. ()
5. ()

13. Avez-vous déjà utilisé un logiciel pour l'évaluation de la qualité de la traduction ?

1. () OUI
2. () NON

14. Si oui, avez-vous compris le résultat de l'évaluation faite par ce programme ?

1. () OUI
2. () NON

15. En cas de réponse positive à la question 8, quel est votre opinion sur ces programmes et ces métriques (indicateurs de médiation de la qualité de la traduction ?

16. Avez-vous un commentaire à faire sur la pertinence des questions et sur le thème de ce questionnaire ?

Merci beaucoup.

ANEXO E – TABULAÇÃO E GRÁFICO DO QUESTIONÁRIO 2

15 réponses

Une méthodologie d'accès et compréhension de textes scientifiques en Portugais par des Étrangers

Recherche.These.Doctorale

f

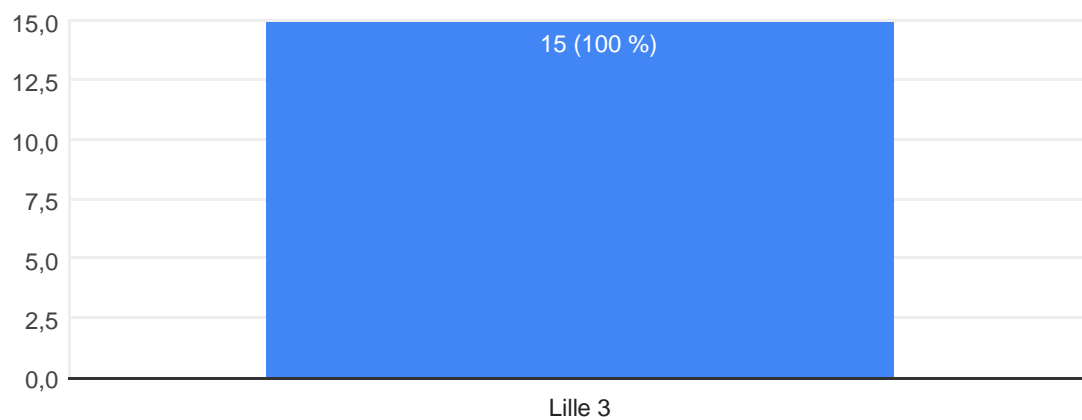
Langue: Français

1. Identification

Nom, Prénom: 15 réponses

L'identification des répondants a été supprimé pour des questions de
privacité.

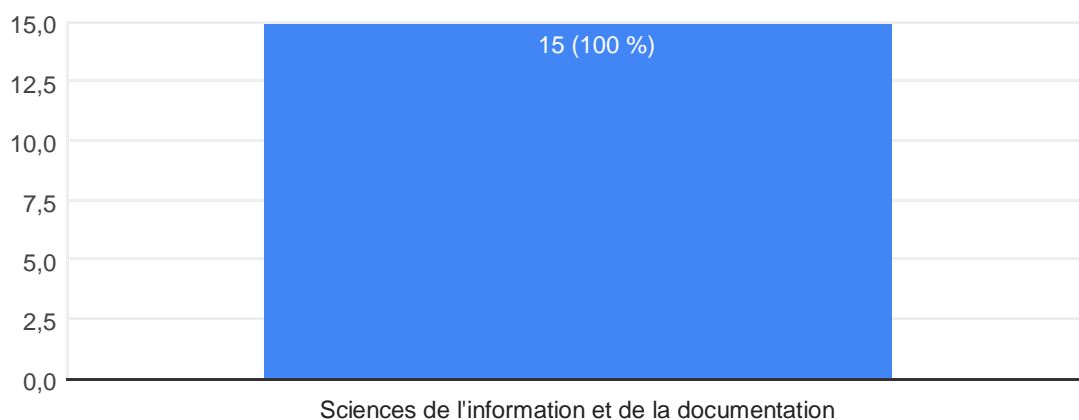
Université: 15 réponses



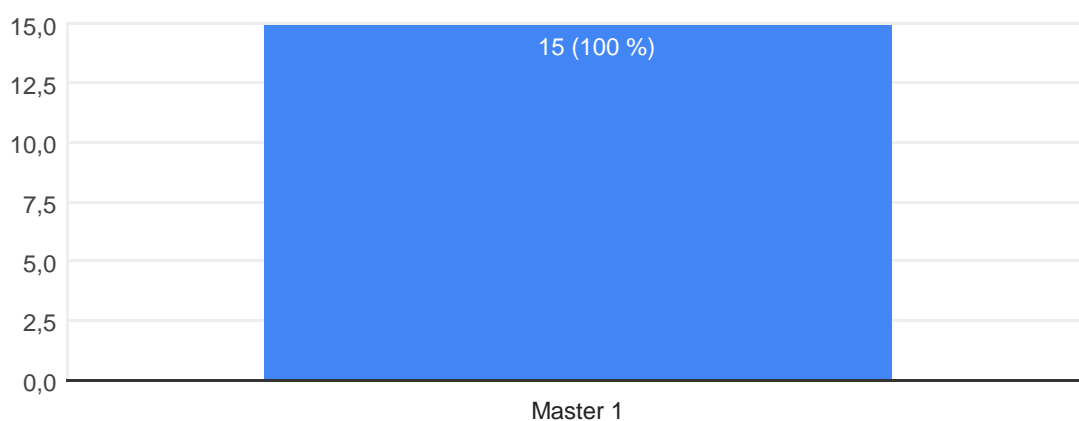
Inscription dans l'Institution: 15 réponses

Ces données ont été supprimées pour des raisons de confidentialité

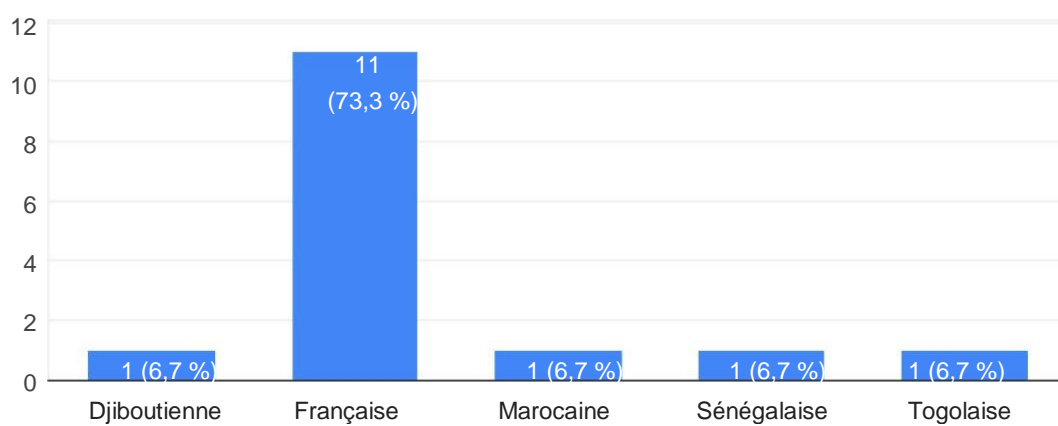
Cours (spécialité): 15 réponses



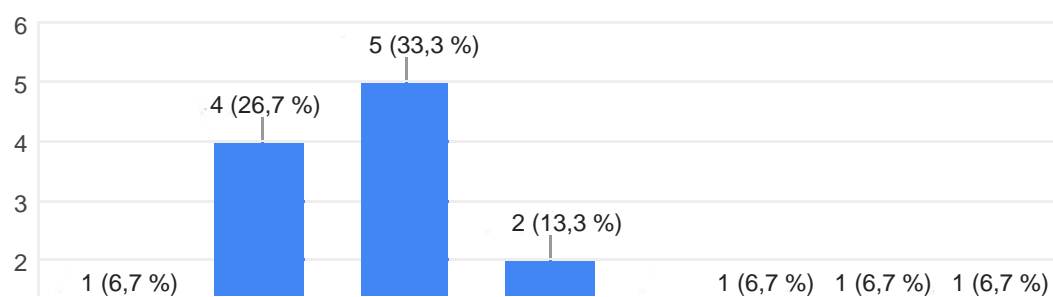
Niveau de la Spécialité (Bac, Master, Doctorat): 15 réponses



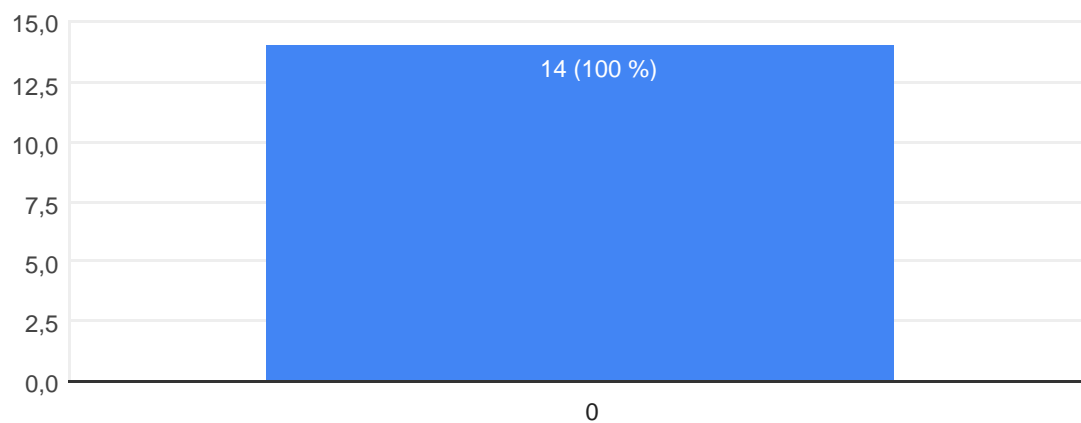
Nationalité: 15 réponses



Age: 15 réponses

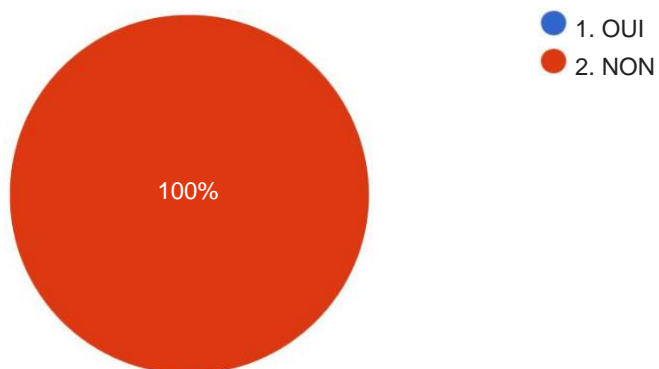


Nombre d'années d'étude du Portugais: 14 réponses

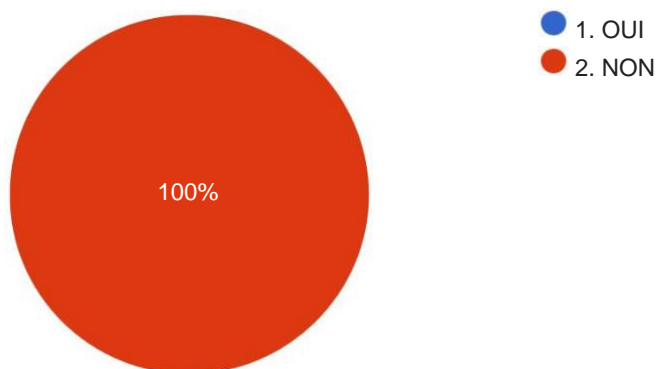


2 – Questions générales

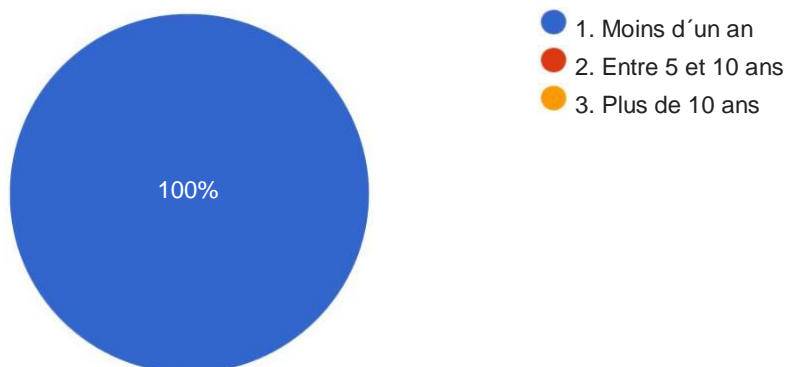
1. Parlez-vous portugais dans votre cercle familial ? 15 réponses



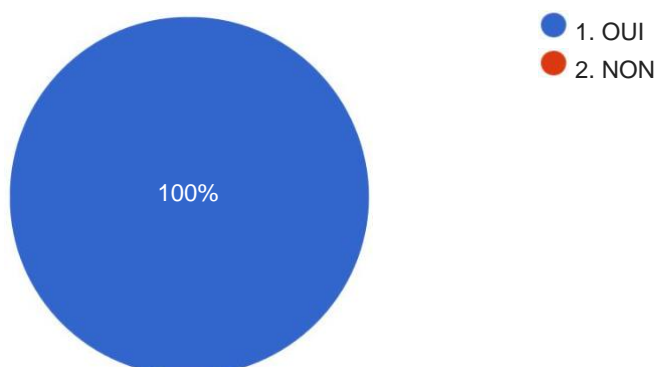
2. Avez vous le diplôme CELPE-BRAS ou un diploma équivalent de profiçence en Portugais 15 réponses



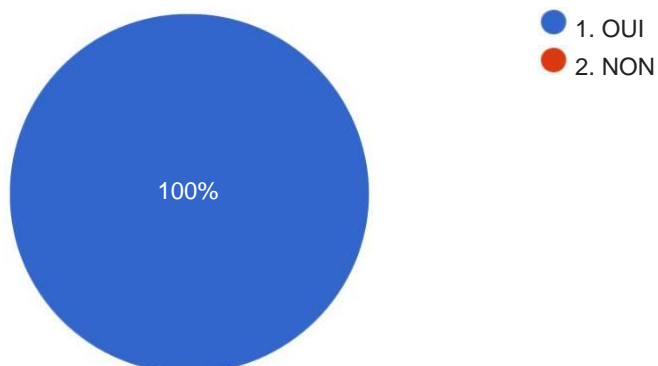
3. Si oui, depuis combien de temps ? Une réponse



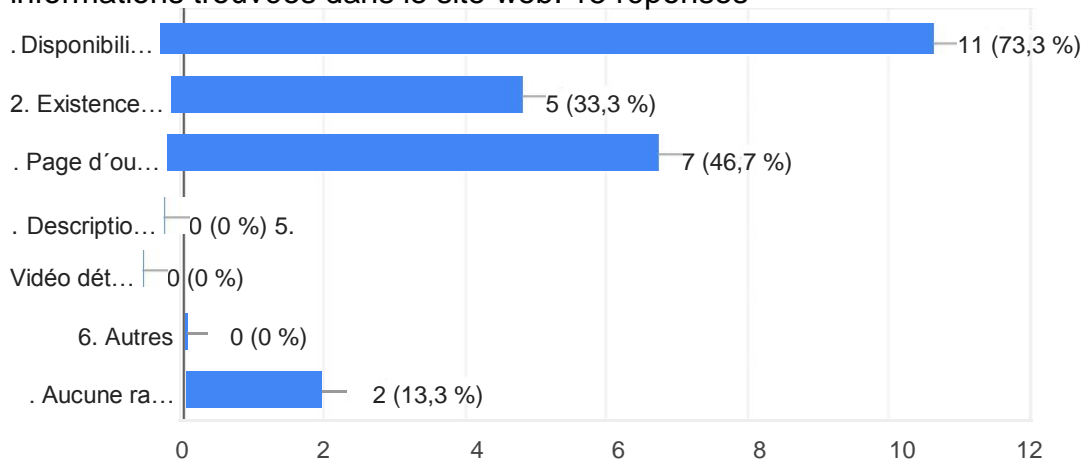
4. Accédez-vous à des sites web en français ? 15 réponses



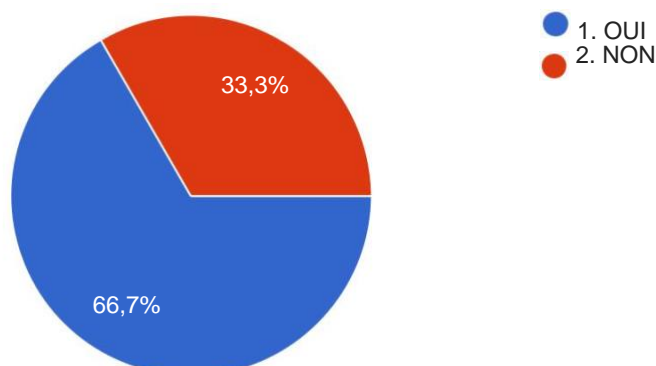
5. Si oui, quand vous accédez à ces sites web, obtenez-vous facilement l'information demandée en français ? 15 réponses



6. Si oui, cochez quelles sont les raisons qui expliquent la facilité de l'accès aux informations trouvées dans le site web: 15 réponses



7. Sur la base de la description des ressources numériques qui ont été présentées dans l'introduction de ce questionnaire (où il est demandé à chacun/une de visiter la BDTD et un deuxième site web), rencontrez-vous des difficultés pour trouver l'information recherchée (thèse, dissertation, article) dans des sites web en Portugais ? 15 réponses



8. Si oui, décrivez brièvement quelles sont ces difficultés 11 réponses Je ne parle pas du tout le portugais donc difficulté de compréhension La barrière de la langue malgré la possibilité d'une recherche en anglais. Je ne parle pas portugais, j'ai donc des problèmes de compréhension
Je ne comprend pas la langue La langue

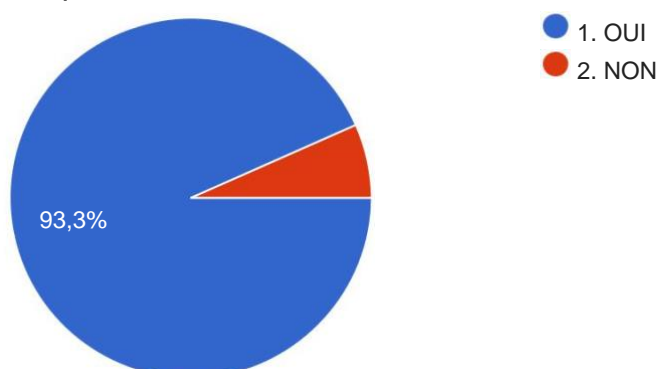
Les titres français ne sont pas retrouvés par la base de données. Difficultés dans l'orientation sur le site, dues à l'incompréhension de la langue (plus particulièrement pour le site SciELO que sur la BDTD). Je ne parle pas le portugais donc je ne comprends pas ce qui est écrit, simplement.

Problème liés à la traduction de la langue vue qu'on a pas des notions ou la maîtrise de la langue portugaise

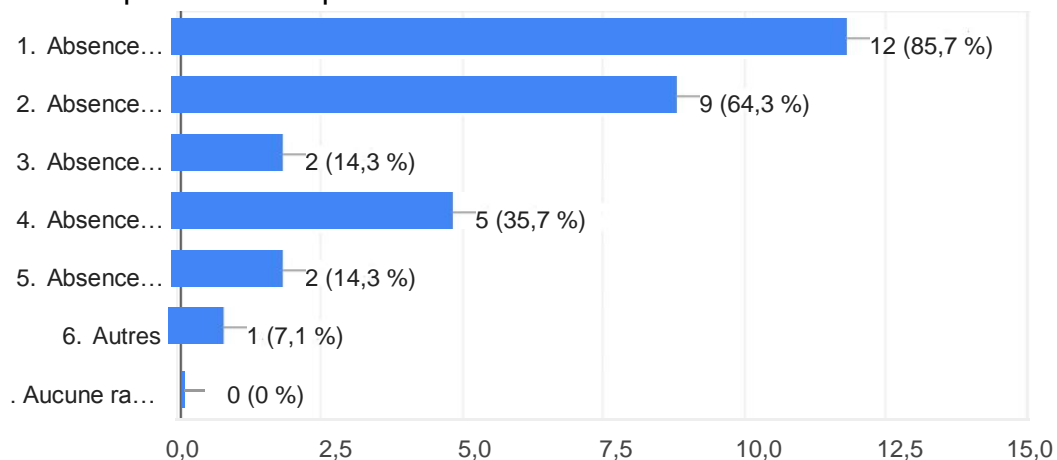
non compréhension du langue portugais

INCOMPRÉHENSION DE LA LANGUE PORTUGAISE

9. Une fois que vous avez trouvé le document recherché (thèse, dissertation, article), avez-vous des difficultés pour comprendre l'information obtenue dans des sites web en Portugais ? 15 réponses

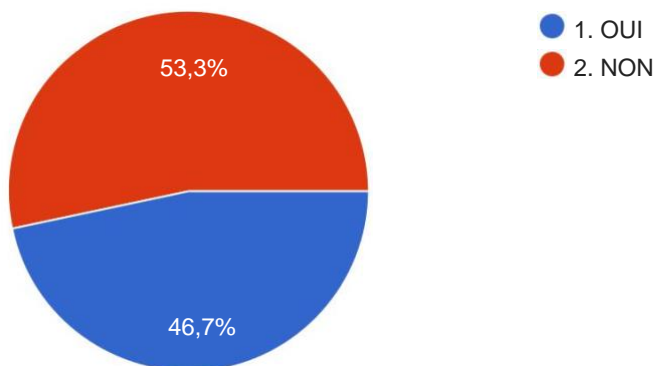


10. Lesquelles ? 14 réponses

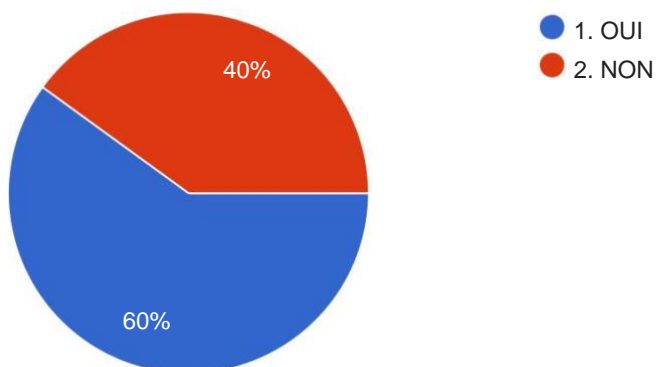


3 – Questions sur les ressources numériques d'information scientifique en Portugais

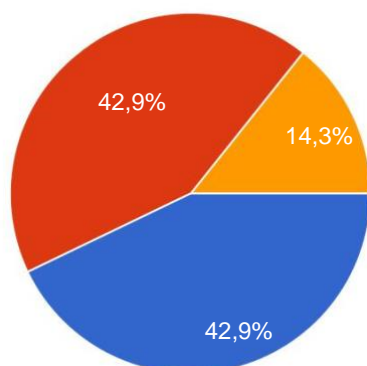
1. Avez-vous consulté un mémoire de master dans la Bibliothèque Numérique de Thèses et Mémoires ? 15 réponses



2. Si oui, avez-vous compris le contenu du Mémoire de Master consulté ? 10 réponses

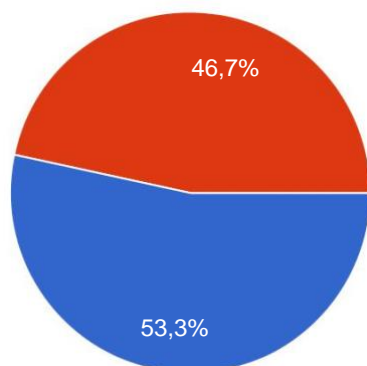


2.1. Si OUI, à quel niveau ? 7 réponses



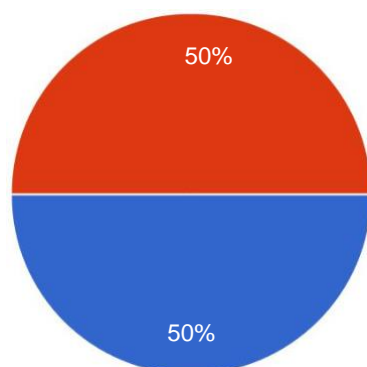
- 1. Simplement le titre du Mémoire de Master
- 2. Le sens de quelques mots
- 3. Uniquement quelques aspects du Mémoire
- 4. La totalité du Mémoire
- 5. Le sommaire du Mémoire

3. Avez-vous consulté une Thèse de Doctorat de la Bibliothèque Numérique de Thèses et Dissertations ? 15 réponses



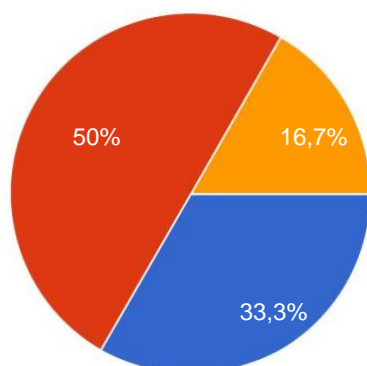
- 1. OUI
- 2. NON

4. Si oui, avez vous compris le contenu de la Thèse consultée ? 10 réponses



- 1. OUI
- 2. NON

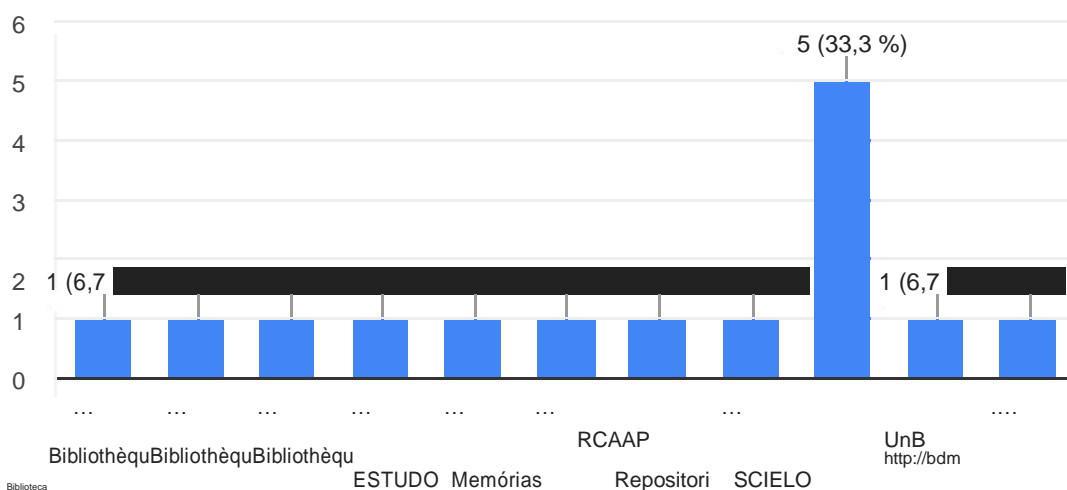
4.1. Si OUI, à quel niveau 6 réponses



- 1. Simplement le titre de la Thèse
- 2. Le sens de quelques mots
- 3. Uniquement quelques aspects de la Thèse
- 4. La totalité de la Thèse
- 5. Le sommaire de la Thèse

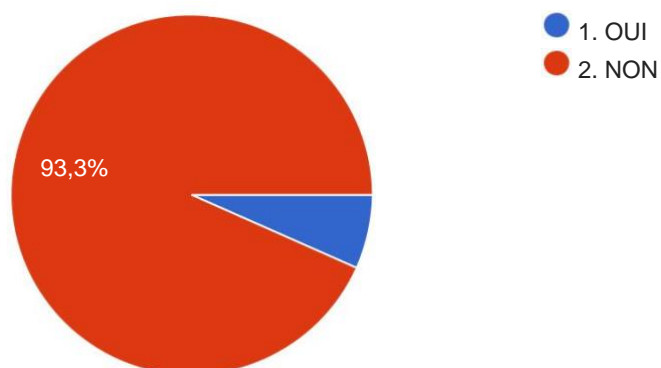
5. Citez le titre de la deuxième ressource numérique que vous avez consulté.

15 réponses

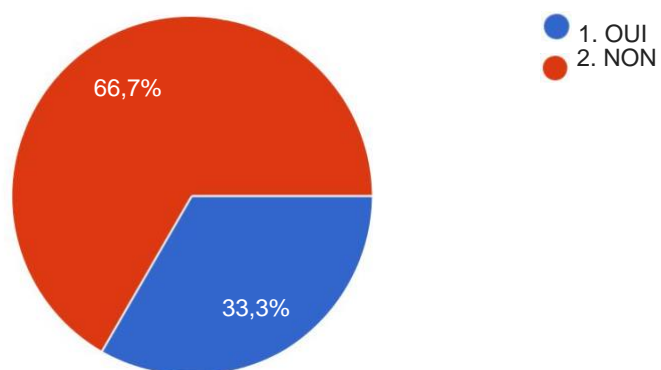


6. Connaissez-vous le Répertoire Institutionnel de la Université de Brasília (UnB) ?

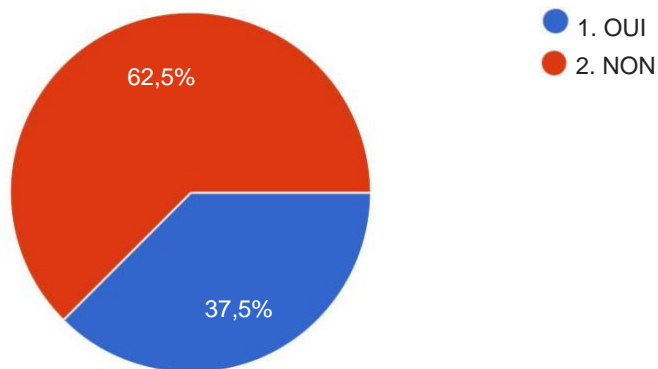
15 réponses



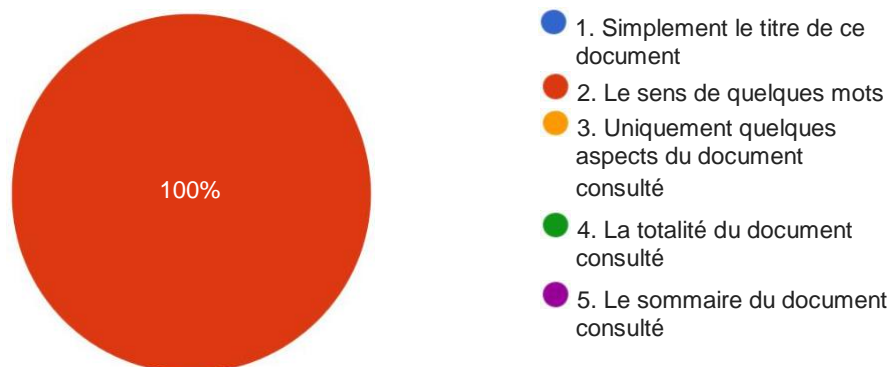
7. Si oui, avez-vous eu accès à au moins un document sur la liste du répertoire institutionnel de l'UnB ? 6 réponses



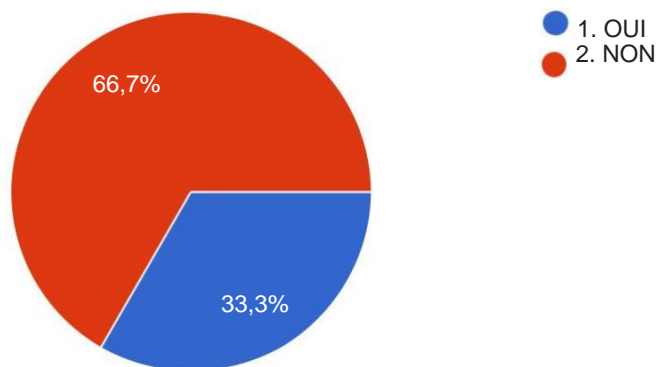
8. Si oui, avez-vous compris le contenu de ce document ? 8 réponses



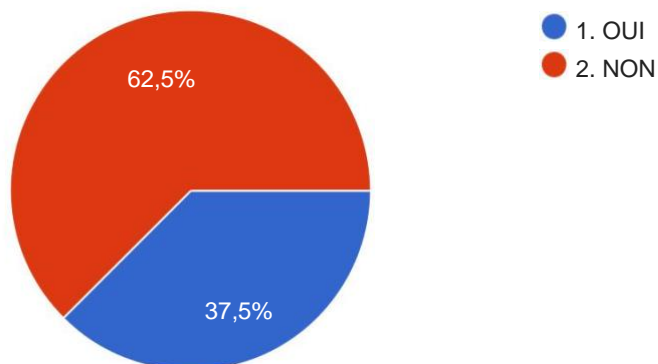
8.1. Se OUI, à quel niveau ? 2 réponses



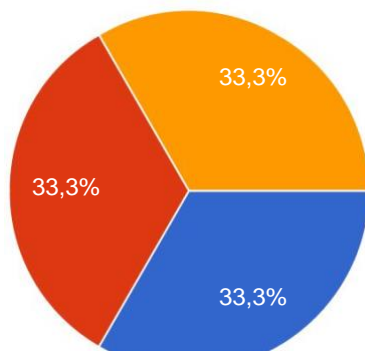
9. Avez-vous consulté des Mémoires de fin de cursus dans la base de données ou dans le répertoire institutionnel de l'UnB ? 15 réponses



10. Si oui, avez-vous compris le contenu du Mémoire de fin de cursus consulté ? 8 réponses

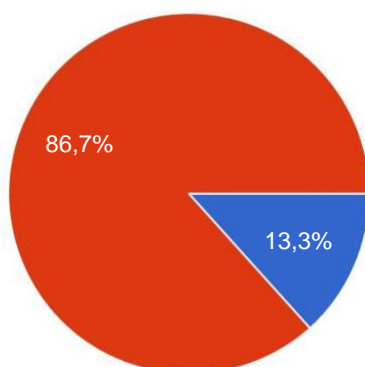


10.1 Se OUI, à quel niveau ? 3 réponses



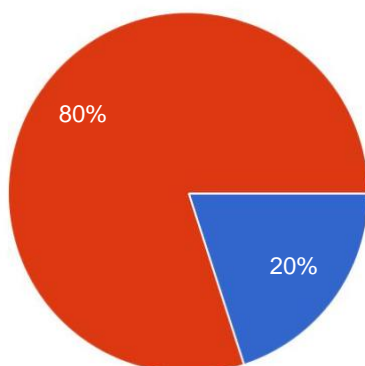
- 1. Simplement le titre de ce Mémoire de fin de cursus
- 2. Le sens de quelques mots
- 3. Uniquement quelques aspects du Mémoire consulté
- 4. La totalité du Mémoire consulté
- 5. Le sommaire du Mémoire consulté

11. Avez-vous consulté un Mémoire de Master disponible sur la liste du répertoire institutionnel de l'UnB ? 15 réponses



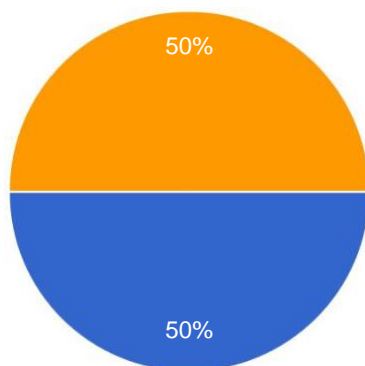
- 1. OUI
- 2. NON

12. Si oui, avez-vous compris le contenu du Mémoire consulté ? 5 réponses



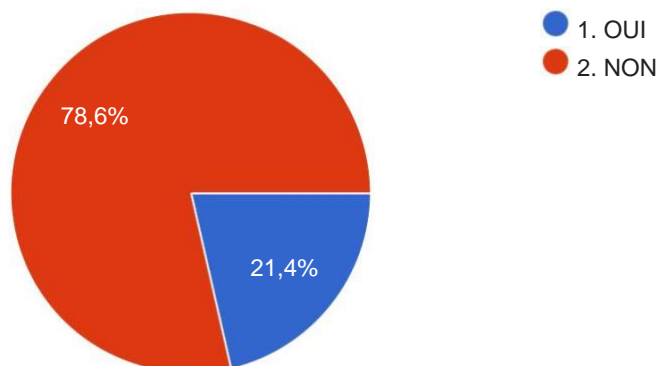
- 1. OUI
- 2. NON

12.1 Si OUI, à quel niveau 2 réponses

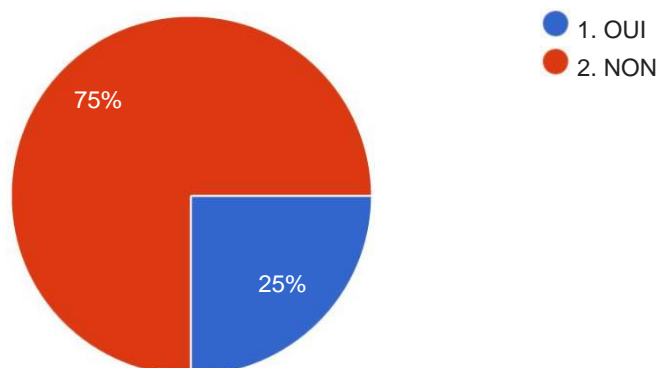


- 1. Simplement le titre de ce Mémoire de Master
- 2. Le sens de quelques mots
- 3. Uniquement quelques aspects du Mémoire consulté
- 4. La totalité du Mémoire de Master consulté
- 5. Le sommaire du Mémoire de Master consulté

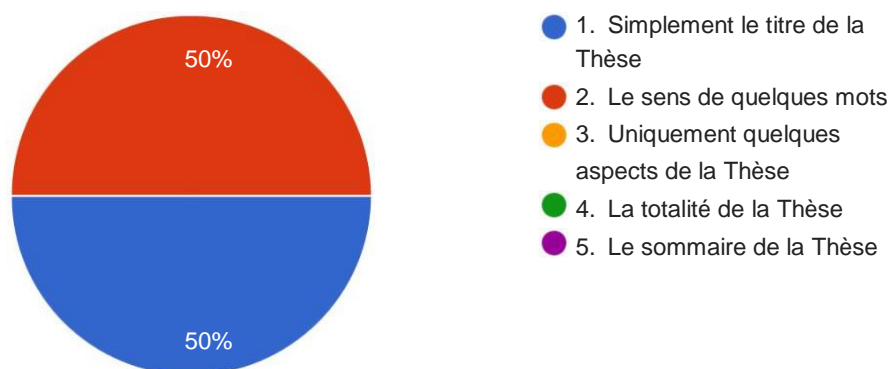
13. Avez-vous consulté une Thèse de Doctorat disponible sur la liste du répertoire institutionnel de l'UnB ? 14 réponses



14. Si oui, avez-vous compris le contenu de la Thèse consultée ? 4 réponses



14.1. Si OUI, à quel niveau ? 2 réponses



15. SVP, indiquez ci-après le titre des deux documents qui vous avez consulté. 15 réponses

Practices in social media in the travel education business, A fan page da Biblioteca Demonstrativa : diretrizes para uso de mídias sociais

DISSERTAÇÃO : GÊNERO OU TIPO TEXTUAL ? / ESTUDO E

DESENVOLVIMENTO DE DISPOSITIVOS ACESSÓRIOS DE POSTES DE CONCRETO ARMADO PARA REDES DE DISTRIBUIÇÃO DE ENERGIA

INFORMATIONAL FRICTIONS AND INFLATION DYNAMICS , O informal/formal na trajetória profissional de dois professores de música popular

Estudo sobre a iconografia de Apis durante o período farainico / Pobres faraos divinos

OPEN ARCHIVES: panorama dos repositórios Um estudo sobre interface cérebro-computador

Instituto de Pesquisas Ambientais em Bacias Hidrográficas

Das materialidades da literatura : a reinvenção da vida e o acervo de narrativas orais urbano-digitais

"AVALIAÇÃO IN VITRO DE MATERIAIS E TÉCNICAS DE TRANSFERÊNCIA EM IMPLANTES MÚLTIPLOS"

"Escola e comunidade : a construção conjunta de valores" de Moreira, Carmen Silvia et Silva, Sonia Aparecida Ignacio et "Família e Suicídio" de Frazão, Pedro Santos, José Carlos Pereira dos Sampaio, Daniel

RECOMMENDATION BASED ON DATA MINING FOR RELATIONSHIP MARKETING

QUEM SÃO OS ESTRANGEIROS EXPULSOS DO BRASIL? Um estudo exploratório sobre o regime jurídico de expulsão de estrangeiros Aprendendo e ensinando : o que pensam os agentes comunitários de saúde sobre sua formação continuada /

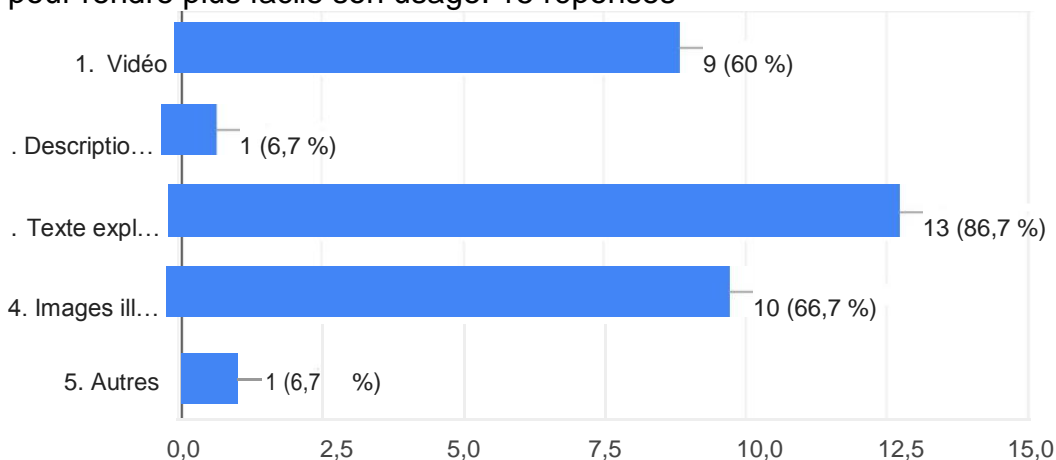
Os desafios da televisão digital - Rosalia Lara de Moura:

Aprendendo e ensinando : o que pensam os agentes comunitários de saúde sobre sua formação continuada

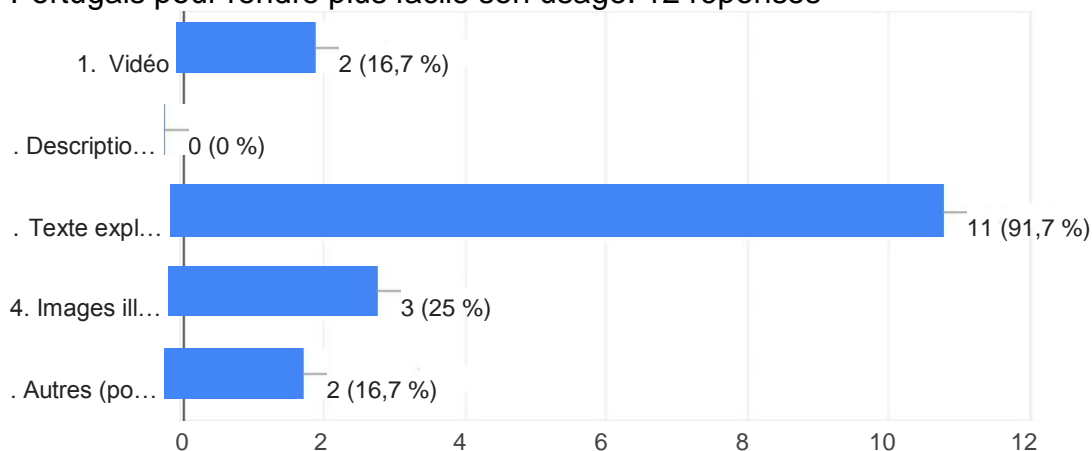
PD Final 12 12 12 Modo Compat.pdf et Abertura comercial, inflação e empreendedorismo de Antonio Nascimento Júnior

4. Questions techniques

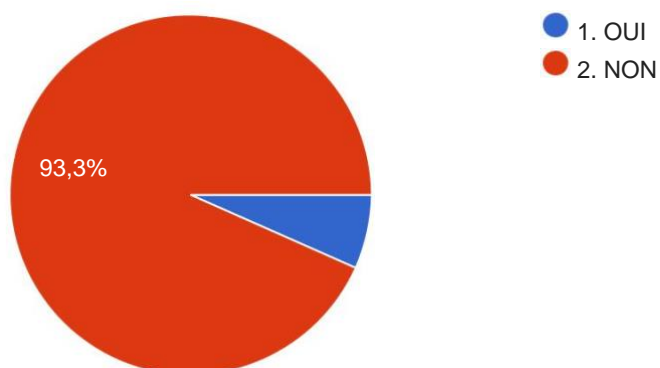
1. Cochez les ressources que vous avez rencontrées dans les sites web en français pour rendre plus facile son usage: 15 réponses



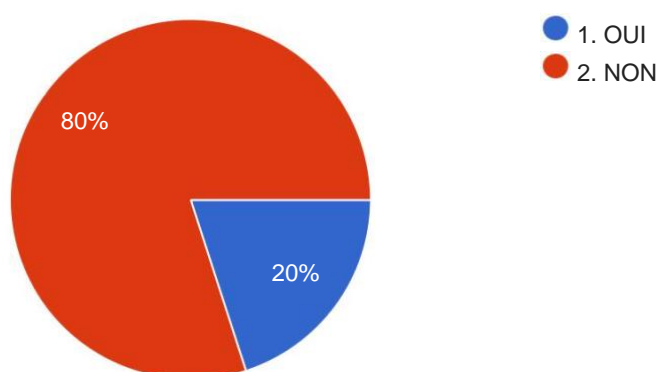
2. Cochez les options que vous avez rencontrées dans les sites web écrits en Portugais pour rendre plus facile son usage: 12 réponses



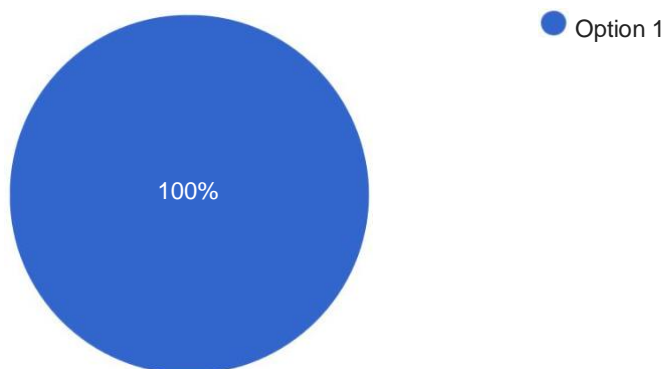
3. Avez-vous déjà utilisé des logiciels de résumé automatique de textes en portugais ? 15 réponses



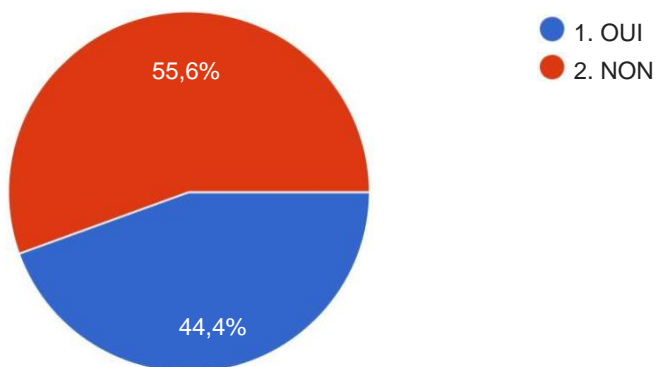
4. Avez-vous déjà utilisé une quelconque ressource informatique (logiciels, dictionnaire ou terminologie électronique, thésaurus informatisé, logiciel de fouille de textes, autres) pour vous aider dans la compréhension de contenus écrits en portugais ? 15 réponses



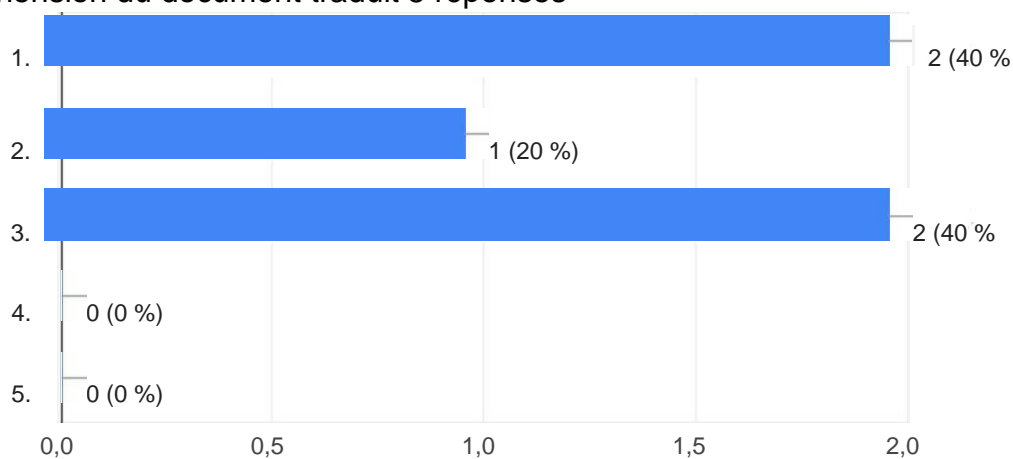
5. Avez-vous déjà utilisé des logiciels de traduction automatique de la langue (TAL) pour traduire des contenus de sites web du portugais vers le français 5 réponses



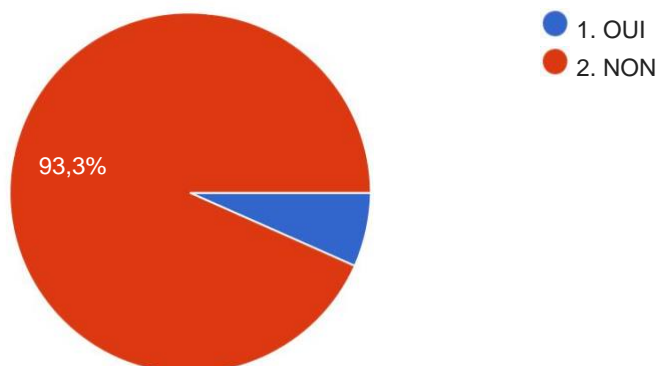
6. Si oui, avez-vous compris le résultat produit par le logiciel de traduction automatique du portugais vers le français ? 9 réponses



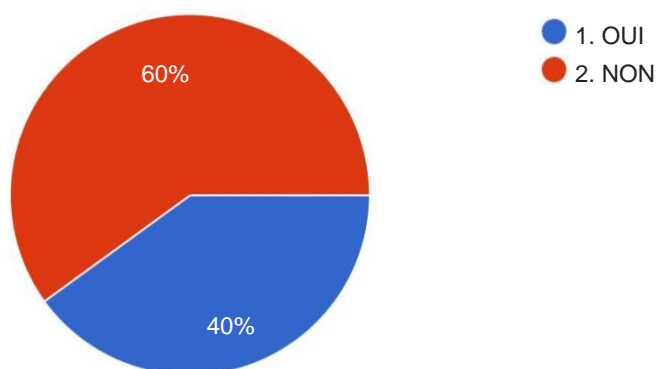
7. Indiquez sur une échelle de 1 a 5 l'utilité de ce logiciel de traduction pour la compréhension du document traduit 5 réponses



8. Avez-vous déjà utilisé un logiciel d'évaluation de la qualité d'une traduction ? 15 réponses



9. Si oui, avez-vous compris le résultat de l'évaluation faite par ce programme ? 5 réponses



10. En cas de réponse positive à la question 8, quel est votre opinion sur ces programmes et ces métriques (indicateurs de médiation de la qualité de la traduction) ? Une réponse

Il y a difficulté de fiabilité dès lors qu'on ne maîtrise pas la langue complètement. Importance de la sémantique.

11. Avez-vous un commentaire à faire sur la pertinence des questions et sur le thème de ce questionnaire ? 5 réponses

Non (2)

Il faudrait parfois changer le "mode" de question. Il n'est pas toujours possible d'effacer une réponse, involontairement cochée au préalable. oui, concernant la redirection automatique des questions ça dépend de la réponse précédente

très intéressant

Merci beaucoup

# Quest	Université:	Cours (spécialité):	Niveau de la Spécialité (Bac, Master, Doctorat):	Nationalité:	Age:	Nombre d'années d'étude du Portugais:
1	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	23	0
2	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	24	0
3	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	24	0
4	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	23	0

# Quest	Université	Cours (spécialité)	Niveau de la Spécialité (Bac, Master, Doctorat)	Nationalité	Age	Nombre d'années d'étude du Portugais
5	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	31	0
6	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	23	0
7	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	24	
8	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	23	0

# Quest	Université	Cours (spécialité)	Niveau de la Spécialité (Bac, Master, Doctorat)	Nationalité	Age	Nombre d'années d'étude du Portugais
9	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	21	0
10	Lille 3	Sciences de l'information et de la documentation	Master 1	Marocaine	25	0
11	Lille 3	Sciences de l'information et de la documentation	Master 1	Togolaise	25	0
12	Lille 3	Sciences de l'information et de la documentation	Master 1	Djiboutienne	24	0

# Quest	Université	Cours (spécialité)	Niveau de la Spécialité (Bac, Master, Doctorat)	Nationalité	Age	Nombre d'années d'étude du Portugais
13	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	46	0
14	Lille 3	Sciences de l'information et de la documentation	Master 1	Sénégalaise	26	0
15	Lille 3	Sciences de l'information et de la documentation	Master 1	Française	24	0

2 - Questions générales									
1. Parlez-vous portugais dans votre cercle familial ?	2. Avez-vous le diplôme CELPE-BRAS ou un diploma équivalent de proficience en Portugais	3. Si oui, depuis combien de temps ?	4. Accédez-vous à des sites web en français ?	5. Si oui, quand vous accédez à ces sites web, obtenez-vous facilement l'information demandée en français ?	6. Si oui, cochez quelles sont les raisons qui expliquent la facilité de l'accès aux informations trouvées dans le site web:	7. Sur la base de la description des ressources numériques qui ont été présentées dans l'introduction de ce questionnaire (où il est demandé à chacun/une de visiter la BDTD et un deuxième site web), rencontrez-vous des difficultés pour trouver l'informati	8. Si oui, décrivez brièvement quelles sont ces difficultés	9. Une fois que vous avez trouvé le document recherché (thèse, dissertation, article), avez-vous des difficultés pour comprendre l'information obtenue dans des sites web en Portugais ?	10. Lesquelles ?
2. NON	2. NON		1. OUI	1. OUI	7. Aucune raison spécifique	1. OUI	Je ne parle pas du tout le portugais donc difficulté de compréhension	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une

									rubrique spécifique et descriptive pour l'accès pour des français
--	--	--	--	--	--	--	--	--	--

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web, 3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;	1. OUI	La barrière de la langue malgré la possibilité d'une recherche en anglais.	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web	1. OUI	Je ne parle pas portugais, j'ai donc des problèmes de compréhension	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français, 4. Absence d'une description orale en français,

									avec des explications sur les fonctionnalités du site web et de son contenu en français;
--	--	--	--	--	--	--	--	--	--

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web	1. OUI	Je ne comprend pas la langue	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web, 2. Existence d'une rubrique spécifique en français pour expliquer le contenu du site web;; 3. Page d'ouverture avec des	2. NON		1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français, 4. Absence d'une description

					illustrations adéquates à l'usage par des français;				orale en français, avec des explications sur les fonctionnalités du site web et de son contenu en français;
--	--	--	--	--	--	--	--	--	--

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web	1. OUI	La langue	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français, 6. Autres
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web	1. OUI	Les titres français ne sont pas retrouvés par la base de données.	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	7. Aucune raison spécifique	1. OUI	Difficultés dans l'orientation sur le site, dues à l'incompréhension de la langue (plus particulièrement pour le site SciELO que sur la BDTD).	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web	1. OUI	Je ne parle pas le portugais donc je ne comprends pas ce qui est écrit, simplement.	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français
2. NON	2. NON	1. Moins d'un an	1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web, 2. Existence d'une rubrique	2. NON		2. NON	

					spécifique en français pour expliquer le contenu du site web;; 3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;				
--	--	--	--	--	---	--	--	--	--

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;	1. OUI	Problème liés à la traduction de la langue vue qu'on a pas des notions ou la maîtrise de la langue portugaise	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web, 2. Existence d'une rubrique spécifique en français pour expliquer le contenu du site web;; 3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;	2. NON	non compréhension du langue portugais	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);, 2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français, 3. Absence de figures illustratives adéquates à l'usage par des français dans la page d'ouverture du site;; 4. Absence d'une

									<p>description orale en français, avec des explications sur les fonctionnalités du site web et de son contenu en français;, 5. Absence d'une vidéo explicative des fonctionnalités du site web et de son contenu en français;</p>
--	--	--	--	--	--	--	--	--	---

2 - Questions générales								
2. NON	2. NON		1. OUI	1. OUI	<p>1. Disponibilité d'un menu pour le choix du français comme langue du site web, 2. Existence d'une rubrique spécifique en français pour expliquer le contenu du site web;, 3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;</p>	2. NON		<p>2. Absence d'une rubrique spécifique et descriptive pour l'accès pour des français, 4. Absence d'une description orale en français, avec des explications sur les fonctionnalités du site web et de son contenu en français;, 5. Absence d'une vidéo explicative des fonctionnalités du site web et de son contenu en français;</p>

2 - Questions générales									
2. NON	2. NON		1. OUI	1. OUI	2. Existence d'une rubrique spécifique en français pour expliquer le contenu du site web;	1. OUI	INCOMPRÉHENSION DE LA LANGUE PORTUGAISE	1. OUI	1. Absence d'un menu pour le choix de la langue d'usage (français, anglais, espagnol);
2. NON	2. NON		1. OUI	1. OUI	1. Disponibilité d'un menu pour le choix du français comme langue du site web, 3. Page d'ouverture avec des illustrations adéquates à l'usage par des français;	2. NON		1. OUI	3. Absence de figures illustratives adéquates à l'usage par des français dans la page d'ouverture du site;, 4. Absence d'une description orale en français, avec des explications sur les fonctionnalités du site web et de son contenu en français;

3. Questions sur les ressources numériques d'information scientifique en Portugais						
1. Avez-vous consulté un mémoire de master dans la Bibliothèque Numérique de Thèses et Mémoires ?	2. Si oui, avez-vous compris le contenu du Mémoire de Master consulté ?	2.1. Si OUI, à quel niveau ?	3. Avez-vous consulté une Thèse de Doctorat de la Bibliothèque Numérique de Thèses et Dissertations ?	4. Si oui, avez vous compris le contenu de la Thèse consultée?	4.1. Si OUI, à quel niveau	5. Citez le titre de la deuxième ressource numérique que vous avez consulté.
2. NON			2. NON			Bibliothèque Numérique de Mémoires de fin de cursus de l'Université de Brasília
1. OUI	1. OUI	2. Le sens de quelques mots	1. OUI	1. OUI	2. Le sens de quelques mots	ESTUDO E DESENVOLVIMENTO DE DISPOSITIVOS ACESSÓRIOS DE POSTES DE CONCRETO ARMADO PARA REDES DE DISTRIBUIÇÃO DE ENERGIA
2. NON			2. NON			Biblioteca digital de monografias
2. NON			1. OUI	1. OUI	2. Le sens de quelques mots	Repositorio Institucional RIUnB
2. NON			2. NON			Bibliothèque Numérique de Mémoires de fin de cursus de l'Université de Brasília, http://bdm.unb.br ;
1. OUI	2. NON		1. OUI	2. NON		Memórias de um escritor bem-comportado : Autran Dourado
1. OUI	2. NON		2. NON			SCIELO
1. OUI	1. OUI	2. Le sens de quelques mots	1. OUI	1. OUI	2. Le sens de quelques mots	SCIELO
1. OUI	1. OUI	1. Simplement le titre du Mémoire de Master	1. OUI	2. NON		RCAAP

2. NON	2. NON	1. Simplement le titre du Mémoire de Master	2. NON			SCIELO
1. OUI	1. OUI	3. Uniquement quelques aspects du Mémoire	2. NON	1. OUI	3. Uniquement quelques aspects de la Thèse	http://bdm.unb.br
1. OUI	2. NON		1. OUI	2. NON		SCIELO
2. NON			2. NON	2. NON		SCIELO
2. NON	1. OUI	2. Le sens de quelques mots	1. OUI	2. NON	1. Simplement le titre de la Thèse	UnB
2. NON	1. OUI	1. Simplement le titre du Mémoire de Master	1. OUI	1. OUI	1. Simplement le titre de la Thèse	Bibliothèque Numérique de Dissertations de Master de l'Université de Brasília

3. Questions sur les ressources numériques d'information scientifique em Portugais						
6. Connaissez-vous le Répertoire Institutionnel de la Université de Brasília (UnB) ?	7. Si oui, avez-vous eu accès à au moins un document sur la liste du répertoire institutionnel de l'UnB ?	8. Si oui, avez-vous compris le contenu de ce document ?	8.1. Se OUI, à quel niveau ?	9. Avez-vous consulté des Mémoires de fin de cursus dans la base de données ou dans le répertoire institutionnel de l'UnB ?	10. Si oui, avez-vous compris le contenu du Mémoire de fin de cursus consulté ?	10.1 Se OUI, à quel niveau ?
2. NON				1. OUI		
2. NON				2. NON		
2. NON				2. NON	2. NON	
2. NON	2. NON	1. OUI	2. Le sens de quelques mots	1. OUI	1. OUI	2. Le sens de quelques mots
1. OUI	1. OUI	2. NON		1. OUI	2. NON	
2. NON				2. NON		
2. NON				2. NON		
2. NON	2. NON	1. OUI		2. NON	2. NON	
2. NON				2. NON		
2. NON				2. NON	1. OUI	1. Simplement le titre de ce Mémoire de fin de cursus
2. NON	2. NON	2. NON		2. NON	2. NON	
2. NON	2. NON	2. NON		2. NON		
2. NON		2. NON		2. NON		
2. NON		2. NON		1. OUI	2. NON	
2. NON	1. OUI	1. OUI	2. Le sens de quelques mots	1. OUI	1. OUI	3. Uniquement quelques aspects du Mémoire consulté

3. Questions sur les ressources numériques d'information scientifique em Portugais						
11. Avez-vous consulté un Mémoire de Master disponible sur la liste du répertoire institutionnel de l'UnB ?	12. Si oui, avez-vous compris le contenu du Mémoire consulté ?	12.1 Si OUI, à quel niveau	13. Avez-vous consulté une Thèse de Doctorat disponible sur la liste du répertoire institutionnel de l'UnB ?	14. Si oui, avez-vous compris le contenu de la Thèse consultée ?	14.1. Si OUI, à quel niveau ?	15. SVP, indiquez ci-après le titre des deux documents qui vous avez consulté.
2. NON			1. OUI			Practices in social media in the travel education business, A fan page da Biblioteca Demonstrativa : diretrizes para uso de mídias sociais
2. NON			2. NON			DISSERTAÇÃO : GÊNERO OU TIPO TEXTUAL ? / ESTUDO E DESENVOLVIMENTO DE DISPOSITIVOS ACESSÓRIOS DE POSTES DE CONCRETO ARMADO PARA REDES DE DISTRIBUIÇÃO DE ENERGIA
2. NON			2. NON			INFORMATIONAL FRICTIONS AND INFLATION DYNAMICS , O informal/formal na trajetória profissional de dois professores de música popular
2. NON			2. NON			Estudo sobre a iconografia de Apis durante o periodo farainico / Pobres faraos divinos
2. NON			2. NON			OPEN ARCHIVES: panorama dos repositórios Um estudo sobre interface cérebro-computador
2. NON			2. NON			Instituto de Pesquisas Ambientais em Bacias Hidrográficas
2. NON			2. NON			Das materialidades da literatura : a reinvenção da vida e o acervo de narrativas orais urbano-digitais
2. NON	2. NON		2. NON	2. NON		"AVALIAÇÃO IN VITRO DE MATERIAIS E TÉCNICAS DE TRANSFERÊNCIA EM IMPLANTES MÚLTIPLOS"
2. NON			2. NON			"Escola e comunidade : a construção conjunta de valores" de Moreira, Carmen Silvia et Silva, Sonia Aparecida Ignacio et "Família e Suicídio" de Frazão, Pedro Santos, José Carlos Pereira dos Sampaio, Daniel

2. NON	2. NON		2. NON			RECOMMENDATION BASED ON DATA MINING FOR RELATIONSHIP MARKETING
2. NON	2. NON		2. NON	2. NON		QUEM SÃO OS ESTRANGEIROS EXPULSOS DO BRASIL? Um estudo exploratório sobre o regime jurídico de expulsão de estrangeiros
2. NON						Aprendendo e ensinando : o que pensam os agentes comunitários de saúde sobre sua formação continuada /
2. NON			2. NON			Os desafios da televisão digital - Rosalia Lara de Moura:
1. OUI	2. NON	1. Simplement le titre de ce Mémoire de Master	1. OUI	2. NON	1. Simplement le titre de la Thèse	Aprendendo e ensinando : o que pensam os agentes comunitários de saúde sobre sua formação continuada
1. OUI	1. OUI	3. Uniquement quelques aspects du Mémoire consulté	1. OUI	1. OUI	2. Le sens de quelques mots	PD Final 12 12 12 Modo Compat.pdf et Abertura comercial, inflação e empreendedorismo de Antonio Nascimento Júnior

4. Questions techniques						
1. Cochez les ressources que vous avez rencontrées dans les sites web en français pour rendre plus facile son usage:	2. Cochez les options que vous avez rencontrées dans les sites web écrits en Portugais pour rendre plus facile son usage:	3. Avez-vous déjà utilisé des logiciels de résumé automatique de textes en portugais ?	4. Avez-vous déjà utilisé une quelconque ressource informatique (logiciels, dictionnaire ou terminologie électronique, thésaurus informatisé, logiciel de fouille de textes, autres) pour vous aider dans la compréhension de contenus écrits en portugais ?	5. Avez-vous déjà utilisé des logiciels de traduction automatique de la langue (TAL) pour traduire des contenus de sites web du portugais vers le français	6. Si oui, avez-vous compris le résultat produit par le logiciel de traduction automatique du portugais vers le français ?	7. Indiquez sur une échelle de 1 à 5 l'utilité de ce logiciel de traduction pour la compréhension du document traduit
3. Texte explicatif (articles, tutoriels), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutoriels), 4. Images illustratives du contenu	2. NON	2. NON			

4. Questions Techniques						
3. Texte explicatif (articles, tutorials)	1. Vidéo, 3. Texte explicatif (articles, tutorials)	2. NON	2. NON			
3. Texte explicatif (articles, tutorials)	1. Vidéo, 3. Texte explicatif (articles, tutorials)	2. NON	2. NON	Option 1	2. NON	1.
1. Vidéo, 3. Texte explicatif (articles, tutorials)	3. Texte explicatif (articles, tutorials)	1. OUI	1. OUI	Option 1	2. NON	2.
1. Vidéo, 4. Images illustratives du contenu		2. NON	2. NON			
3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	2. NON	2. NON		2. NON	
1. Vidéo, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu		2. NON	2. NON			
1. Vidéo, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials)	2. NON	2. NON		1. OUI	
1. Vidéo, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials)	2. NON	2. NON		2. NON	
3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu, 5. Autres (podcasts, musique)	2. NON	1. OUI	Option 1	2. NON	

4. Questions Techniques						
3. Texte explicatif (articles, tutorials)	3. Texte explicatif (articles, tutorials)	2. NON	2. NON	Option 1	1. OUI	3.
1. Vidéo, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials)	2. NON	1. OUI	Option 1	1. OUI	3.
1. Vidéo, 2. Description orale, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu, 5. Autres		2. NON	2. NON			
1. Vidéo	5. Autres (podcasts, musique)	2. NON	2. NON		1. OUI	1.
1. Vidéo, 3. Texte explicatif (articles, tutorials), 4. Images illustratives du contenu	3. Texte explicatif (articles, tutorials)	2. NON	2. NON			

4. Questions Techniques			
8. Avez-vous déjà utilisé un logiciel d'évaluation de la qualité d'une traduction ?	9. Si oui, avez-vous compris le résultat de l'évaluation faite par ce programme ?	10. En cas de réponse positive à la question 8, quel est votre opinion sur ces programmes et ces métriques (indicateurs de médiation de la qualité de la traduction) ?	11. Avez-vous un commentaire à faire sur la pertinence des questions et sur le thème de ce questionnaire ?
2. NON			
2. NON			Il faudrait parfois changer le "mode" de question. Il n'est pas toujours possible d'effacer une réponse, involontairement cochée au préalable.
2. NON			
2. NON			
2. NON			
2. NON			Non
2. NON	2. NON		
2. NON	2. NON		
2. NON	2. NON		oui, concernant la redirection automatique des questions ça dépend de la réponse précédente
4. Questions Techniques			
2. NON	1. OUI		Non
2. NON			
2. NON			
2. NON			très intéressant

1. OUI	1. OUI	Il y a difficulté de fiabilité dès lors qu'on ne maîtrise pas la langue complètement. Importance de la sémantique.	
--------	--------	--	--

ANEXO G – TUTORIAL PARA USO DO APLICATIVO COMPUTACIONAL

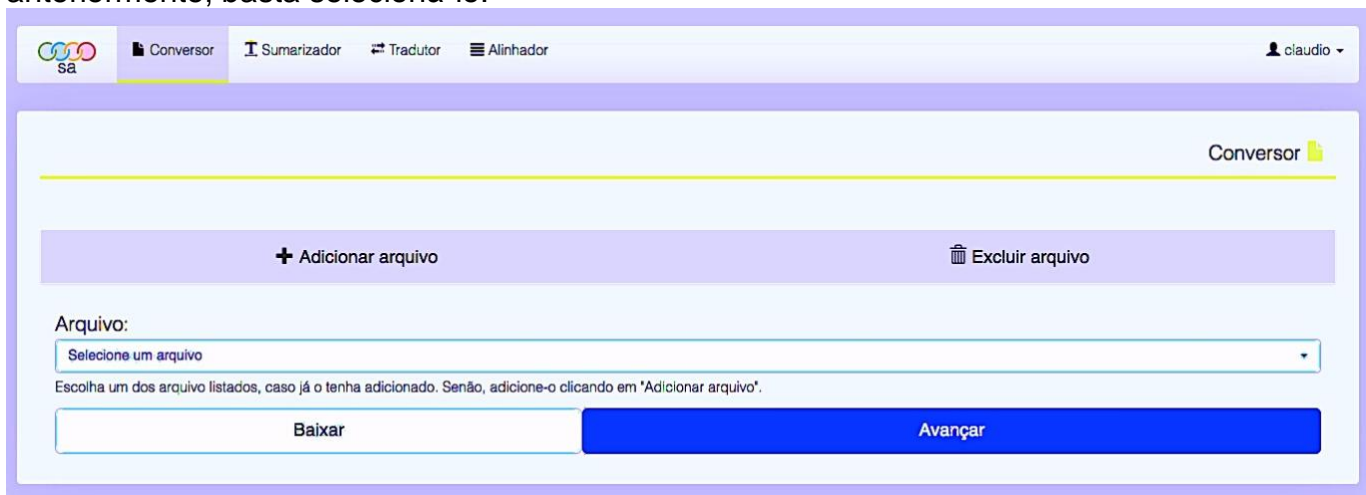
O acesso ao aplicativo está disponível na URL <http://164.41.62.101>.

No primeiro acesso, o usuário deverá alimentar os parâmetros de abertura de conta (nome do usuário, senha). Nos acessos subsequentes, basta usar o login dessa conta.

Ao acessar o sistema, o aplicativo mostra a tela com os 4 componentes a serem usados: conversor, sumarizador, tradutor e alinhador.

Passo 1: Conversão

Transforma o texto fonte em PDF em um texto no formato txt. Se for um arquivo novo, clicar em + Adicionar arquivo para carregá-lo no sistema. Caso o arquivo já tenha sido incluído anteriormente, basta selecioná-lo.



Passo 2: Sumarização

A partir de parâmetros fornecidos pelo usuário (taxa de compressão, início e final do texto a ser sumarizado), produz um texto comprimido de menor tamanho com as sentenças mais relevantes do original.



Passo 3: Tradução automática

Permite escolher a língua alvo do texto a ser traduzido. As opções são: Português, Inglês, Español, Alemão, Francês e Japonês. O resultado será uma tradução do sumário automático obtido na etapa anterior.

The screenshot shows the 'Tradutor' interface. At the top, there is a navigation bar with icons for 'Conversor', 'Sumarizador', 'Tradutor', and 'Alinhador'. The 'Tradutor' tab is selected. Below the navigation bar, there is a header area with the text 'Tradutor' and a small icon. The main content area displays a file name 'tese.computacao' followed by a document icon and '2017_JoséLuisSonccoÁlvarez.txt'. Below this, there is a dropdown menu labeled 'Idioma da tradução:' with 'Francês' selected. At the bottom of the interface, there is a large blue button labeled 'Traduzir'.

Passo 4 – Alinhamento automático

Permite gerar um bi-texto, com as sentenças em português e na língua alvo. O texto assim construído pode ser usado em qualquer biblioteca que se interesse pela implantação do serviço para estrangeiros.

The screenshot shows the 'Alinhador' interface. At the top, there is a navigation bar with icons for 'Conversor', 'Sumarizador', 'Tradutor', and 'Alinhador'. The 'Alinhador' tab is selected. Below the navigation bar, there is a header area with the text 'Alinhador' and a small icon. The main content area displays the same file name 'tese.computacao' followed by a document icon and '2017_JoséLuisSonccoÁlvarez.txt'. Below this, there is a section labeled 'Arquivos para alinhamento sentencial:' with two input fields: 'Sumarizado (Português)' and 'Traduzido (Francês)'. At the bottom of the interface, there is a large blue button labeled 'Alinhar'.

ANEXO H – BI TEXTO DE DISSERTAÇÃO DE MESTRADO



Universidade de Brasília
Biblioteca Central

Serviço de Acesso a Conteúdos Digitais em Português para Estudantes Estrangeiros

Título do Documento Original: Norma linguística e oralidade fingida na tradução de Persépolis

Tipo do Documento: Tese de Mestrado

Lingua do Original: Português

Língua Alvo: Francês

Nome do Arquivo: Claudio.Formulario.Unb.Tese.doc

Sentença Fonte

De certa forma, há uma ampliação dos recursos que serão aplicados na tradução: o autor utilizou a imagem que deve ser considerada pelo tradutor como chave de leitura. |||

A partir dos elementos presentes na linguagem dos quadrinhos, o trabalho do tradutor dessas histórias não se restringe ao texto escrito é preciso levar em conta outros desafios que vão desde uma linguagem icônica até as várias normas linguísticas incidentes no texto. |||

No gênero quadrinhos, os balões expressam a fala de cada personagem e os autores usam os recursos gráficos para expressar a oralidade na escrita, o que leva o pesquisador (e, por conseguinte, o tradutor) a rever os conceitos de língua falada e língua escrita bem como suas inter-relações. |||

Sentença Alvo

D'une certaine façon, il y a une augmentation des ressources qui seront appliquées dans la traduction: l'auteur a utilisé l'image qui doit être considérée par le traducteur comme une clé de lecture.

À partir des éléments présents dans la langue de la bande dessinée, le travail du traducteur de ces histoires ne se limite pas au texte écrit, il faut tenir compte d'autres défis allant d'un langage iconique aux différentes normes linguistiques incidentes dans le texte.

Dans le genre comique, les ballons expriment le discours de chaque personnage et les auteurs utilisent les ressources graphiques pour exprimer l'oralité par écrit, ce qui conduit le chercheur (et donc le traducteur) à examiner les concepts de langue parlée et de langue écrite ainsi que leurs interrelations.

A relação entre língua falada/escrita é muito estreita e a linguagem dos quadrinhos oferece uma característica essencial a mais: a oralidade fingida. |||

 Nos textos das HQ estão presentes recursos da oralidade fingida que são uma tentativa de representar a fala espontânea por meio da linguagem escrita. |||

 O uso abundante de diálogos tem uma estreita relação com a oralidade fingida e as normas que incidem sobre o texto traduzido. |||

 Analisaremos as representações de normas presentes no RG por meio de exemplos de oralidade fingida com vistas à análise linguística das normas que incidiram sobre o texto final traduzido, dando ênfase aos traços gramaticais do português brasileiro (daqui em diante PB): demonstrativos esse/este, formas do imperativo, relações pronominais você/te, emprego dos verbos ter e haver, emprego de nós e a gente. |||

----- •
 Analisar a linguagem verbal/não verbal e a representação de marcas de normas linguísticas presentes na oralidade fingida no texto final traduzido de Persépolis para o PB. |||

 Pretendemos com essa pesquisa responder às perguntas: que representações de normas linguísticas do português brasileiro estão presentes no texto final traduzido de Persépolis? |||

 Como a oralidade fingida foi considerada para a tradução dos balões que representam nas HQ a fala de cada personagem? |||

La relation entre la langue parlée et la langue écrite est très proche et la langue des bandes dessinées offre une caractéristique essentielle: l'oralité prétendue.

Dans les textes de l'HQ, il y a des caractéristiques de l'oralité prétendue qui tentent de représenter le discours spontané par la langue écrite.

L'utilisation abondante des dialogues a une relation étroite avec l'oralité prétendue et les normes qui affectent le texte traduit.

Nous analyserons les représentations des normes présentes dans le GR au moyen d'exemples d'oralité prétendue en vue de l'analyse linguistique des normes axées sur le texte final traduit, soulignant les caractéristiques grammaticales du portugais brésilien (dorénavant PB) les formes des relations impératives et pronominales que vous / te, l'emploi des verbes à avoir et à être, l'emploi de nous et des personnes.

Analyser le langage verbal / non verbal et la représentation des marques linguistiques dans l'oralité prétendue dans le texte final traduit de Persépolis à PB.

Nous avons l'intention avec cette recherche de répondre aux questions: quelles sont les représentations des normes de langue portugaise brésilienne dans le texte final traduit de Persépolis?

Comment l'oralité prétendue a été considérée pour la traduction des ballons qui représentent dans le QG le discours de chaque personnage?

 Fazemos também um levantamento comparativo do panorama descrito pelos especialistas – gramáticos e linguistas – sobre as representações das normas nas ocorrências dos fatos gramaticais do RG Persépolis selecionados para essa pesquisa. |||

 Cirne (2000) explica que as histórias em quadrinhos (daqui em diante HQ), em seus primórdios, foram consideradas por muitos estudiosos e psicólogos textos nocivos à formação dos jovens, histórias escritas sem nenhuma expressividade artística ou textual. |||

 Esse novo tipo de texto, considerado confuso, que era o novo suporte dos quadrinhos, foi apresentado ao público pelos especialistas da época como uma mídia popular de pouca importância e sem nenhum atrativo intelectual. |||

 Para muitos psicólogos, os malefícios da leitura de quadrinhos eram surpreendentemente maiores do que a total ausência de leitura, já que viciava o cérebro em uma “estranha e limitada” forma de escrever, fato explicado por Anselmo: Durante muito tempo as HQ, apontadas como prejudiciais ao desenvolvimento intelectual das crianças, sem qualquer fundamento científico, foram somente objeto de estudos de cunho histórico e artístico. |||

 Com a linguagem oral mais estudada e valorizada, os textos de diálogos, tão utilizados em quadrinhos, são enriquecidos por estruturas mais próximas da língua falada caracterizando personagens e proporcionando variações

Nous faisons également un sondage comparatif sur le panorama décrit par les spécialistes - grammairiens et linguistes - sur les représentations des normes dans les occurrences des faits grammaticaux de RG Persépolis sélectionnés pour cette recherche.

Cirne (2000) explique que les bandes dessinées (ci-après, QG), au début, ont été considérées par de nombreux chercheurs et psychologues, des textes nuisibles à la formation des jeunes, des histoires écrites sans aucune expression artistique ou textuelle.

Ce nouveau type de texte, considéré comme confus, qui était le nouveau soutien de la bande dessinée, a été présenté au public par les spécialistes de l'époque en tant que média populaire de peu d'importance et sans aucune attraction intellectuelle.

Pour de nombreux psychologues, les méfaits de la lecture de la bande dessinée étaient étonnamment supérieurs à l'absence totale de lecture, puisqu'elle viciait le cerveau dans une forme d'écriture "étrange et limitée", un fait expliqué par Anselmo: depuis longtemps, les bandes dessinées, nuisibles au développement intellectuel des enfants, sans aucune base scientifique, ne faisaient que l'objet d'études historiques et artistiques.

dos universos linguísticos aos seus leitores. |||
 Avec le langage oral le plus étudié et le plus apprécié, les textes de dialogue, utilisés dans les bandes dessinées, sont enrichis par des structures plus proches de la langue parlée caractérisant les

personnages et fournissant des variations

 possível reconhecer que, no início, principalmente na era de ouro dos quadrinhos, os anos 1930, surgiram clássicos do gênero como Flash Gordon de Alex Raymond que tem em sua linguagem um delineamento narrativo,

envolvendo “pranchas”, conceito gráfico assim explicado por Mota: A prancha – a página desenhada – é hoje também uma unidade de significação desta forma de linguagem. |||

des univers linguistiques à leurs lecteurs.

il est possible de reconnaître que, au début, principalement dans l'âge d'or des bandes dessinées, les années 1930, les classiques du genre sont apparus comme Flash Gordon d'Alex Raymond qui a dans sa langue une délimitation narrative impliquant des «planches», concept graphique expliqué par Mota: La planche - la page dessinée - est aujourd'hui aussi une unité de sens de cette forme de langue.

ANEXO I – BI-TEXTO DE UMA TESE DE DOUTORADO



Universidade de Brasília
Biblioteca Central

Serviço de Acesso a Conteúdos Digitais em Português para Estudantes Estrangeiros

Título do Documento Original: Cálculo da Distância de Reversão e Construção de Árvores Filogenéticas usando a Ordem dos Genes

Tipo do Documento: Tese de Doutorado

Lingua do Original: Português

Língua Alvo: Francês

Nome do Arquivo: 2017_JoséLuisSonccoÁlvarez_ER1xnU0_a (1)-8.docx

Sentença Fonte

A complexidade deste problema pode variar dependendo se os genes foram abstraídos considerando a sua orientação, gerando permutações com sinal ou não. |||

O problema de ordenação por reversões (usando permutações sem sinal) é um problema de otimização, onde o objetivo é minimizar o número de reversões para transformar um organismo em outro. |||

Os genomas dos organismos são representados como uma sequência de números naturais diferentes, onde cada número representa a ordem de um gene dentro do genoma. |||

A sequência mostrada representa uma solução ótima com só 4 reversões que é a distância de reversão entre estes dois organismos. |||

Sentença Alvo

La complexité de ce problème peut varier selon que les gènes ont été abstraits compte tenu de leur orientation, générant des permutations avec ou sans signe.

Le problème de la commande inverse (en utilisant des permutations non signées) est un problème d'optimisation, où l'objectif est de minimiser le nombre d'inversions pour transformer un organisme en un autre

Les génomes des organismes sont représentés comme une séquence de différents nombres naturels, où chaque nombre représente l'ordre d'un gène dans le génome

La séquence montrée représente une solution optimale avec seulement 4 inversions qui est la distance de réversion entre ces deux organismes



Universidade de Brasília
Biblioteca Central

Por outro lado, o rearranjo de genomas é baseado na análise das ordens dos genes, normalmente representadas como permutações com ou sem sinal. |||

D'autre part, le réarrangement des génomes repose sur l'analyse des ordres des gènes, généralement représentés comme des permutations avec ou sans signal

Nesse contexto, a reconstrução de árvores filogenéticas está relacionada ao problema de rearranjo de múltiplos genomas, cujo objetivo é construir uma árvore filogenética que minimize o custo total da árvore com respeito a uma métrica (e.g. distância de reversão). |||

Dans ce contexte, la reconstruction des arbres phylogénétiques est liée au problème du réarrangement des génomes multiples, dont l'objectif est de construire un arbre phylogénétique qui minimise le coût total de l'arbre par rapport à une métrique (par exemple, la distance d'inversion)

A Figura 1.2 mostra a árvore filogenética do dataset Campanulaceae, a qual contém genomas circulares, esta árvore foi encontrada pelo *software* MGR [19] com um custo total (escore) de 65 2 reversões. |||

La figure 1.2 montre l'arbre phylogénétique de l'ensemble de données Campanulaceae, qui contient des génomes circulants, cet arbre a été trouvé par le logiciel MGR [19] avec un coût total (score) de 65 2 retournements

Hannenhalli e Pezner [44] propuseram o primeiro algoritmo exato de tempo polinomial para resolver o problema de ordenação por reversões de permutações (com sinal). |||

Hannenhalli et Pezner [44] ont proposé le premier algorithme de temps polynomial exact pour résoudre le problème de la commande par inversions de permutations (avec signe)

Depois, como uma extensão natural dessa pesquisa todos esses resultados foram incluídos dentro de um contexto maior que é o da reconstrução de árvores filogenéticas. |||

Ensuite, en tant que prolongement naturel de cette recherche, tous ces résultats ont été inclus dans un contexte plus large qui est la reconstruction des arbres phylogénétiques.



**Universidade de Brasília
Biblioteca Central**

O cálculo da distância de reversão foi incluído no software de filogenia GRAPPA que só usava a distância de pontos de quebra, estendendo dessa maneira a análise filogenética usando uma distância mais significativa desde o ponto de vista evolutivo. |||

.-----
A versão do problema de ordenação por reversões usando permutações sem sinal foi mostrada que é NP-Difícil [21], portanto esta versão do problema é também interessante desde o ponto de vista computacional e combinatório. |||

.-----
Depois foram propostas melhorias sobre a abordagem de Auyeung e Abraham como a inclusão de heurísticas usadas por algoritmos de aproximação [76]. |||

.-----
A maioria das abordagens (e.g. GRAPPA) para solucionar este problema primeiro solucionavam o problema da mediana de 3 genomas, este problema foi demonstrado ser NP-Difícil para diversos modelos evolutivos (reversões, operações DCJ). |||

.-----
Até onde sabemos não foram utilizadas heurísticas para explorar o espaço de soluções quando são usados dados baseados na ordem dos genes. |||

Le calcul de la distance de réversion a été inclus dans le logiciel GRAPPA, qui n'a utilisé que la distance du point de rupture, prolongeant ainsi l'analyse phylogénétique en utilisant une distance plus évolutive

La version du problème de la commande par inversions utilisant des permutations non signées a été révélée NP-Difficult [21], donc cette version du problème est également intéressante d'un point de vue computationnel et combinatoire

Ensuite, des améliorations ont été proposées sur l'approche Auyeung et Abraham comme inclusion des heuristiques utilisées par les algorithmes d'approximation [76]

La plupart des approches (par exemple, GRAPPA) pour résoudre ce problème ont d'abord résolu le problème de la médiane de 3 génomes, ce problème s'est révélé NP-Difficile pour plusieurs modèles évolutifs (inversions, opérations DCJ)

À notre connaissance, les heuristiques n'ont pas été utilisées pour explorer l'espace de la solution lorsque des données basées sur l'ordre des gènes sont utilisées



Universidade de Brasília
Biblioteca Central

Portanto, também precisavam ser propostas abordagens heurísticas para explorar o espaço de soluções de estruturas de árvores no caso de dados baseados na ordem dos genes, e usando como métricas (para avaliar o custo das árvores) as distâncias de reversão e DCJ. |||

Propuseram-se diversos algoritmos evolutivos para o problema de ordenação por reversões usando dados baseados na ordem dos genes (permutações sem sinal). |||

Desenvolveu-se um software baseado em heurísticas para a reconstrução de árvores filogenéticas usando como entrada dados baseados na ordem dos genes (permutações com sinal). |||

Foram obtidos os seguintes resultados específicos: • Desenvolvimento de novos algoritmos baseados no algoritmo genético (proposto em [76]) para problema de ordenação por reversões, o qual é aprimorado usando outras heurísticas como busca local, e busca por oposição (Referências: [77]). |||

----- •
Realizaram-se experimentos utilizando como entrada dados baseados na ordem dos genes, representados neste caso como permutações sem sinal as quais foram geradas de diversas formas: de forma aleatória, ou baseadas em dados biológicos (Referências: [79, 74, 77]). ||| •

Par conséquent, les approches heuristiques pour explorer l'espace des solutions des structures arborescentes dans le cas des données basées sur l'ordre des gènes et l'utilisation de métriques (pour évaluer le coût des arbres), les distances de réversion et les MCJ ont également dû être proposées

Plusieurs algorithmes d'évolution ont été proposés pour le problème de la commande par inversions en utilisant des données basées sur l'ordre des gènes (permutations non signées)

Le logiciel basé sur l'heuristique a été développé pour la reconstruction d'arbres phylogénétiques en utilisant des données d'entrée basées sur l'ordre des gènes (permutations signées)

Les résultats spécifiques suivants ont été obtenus: • Développement de nouveaux algorithmes basés sur l'algorithme génétique (proposé dans [76]) pour le problème de l'ordre par inversions, amélioré en utilisant d'autres heuristiques comme recherche locale et recherche d'opposition (Références: [77])

Les expériences ont été effectuées en tant que données d'entrée en fonction de l'ordre des gènes, représentés dans ce cas en tant que permutations non signées qui ont été générées de plusieurs façons: au hasard ou sur la base de données biologiques (Références: [79, 74, 77])



**Universidade de Brasília
Biblioteca Central**

Realizou-se uma comparação estatística dos resultados dos experimentos para determinar qual é o melhor algoritmo e se os resultados deste algoritmo tem uma diferença estatisticamente significativa com respeito aos outros algoritmos. [51], para o problema da pequena filogenia tomando como dados de entrada um conjunto de genomas baseados na ordem dos genes, e uma estrutura de uma árvore (Referências: [78]). ||| •

Une comparaison statistique des résultats des expériences a été effectuée pour déterminer le meilleur algorithme et si les résultats de cet algorithme ont une différence statistiquement significative par rapport aux autres algorithmes.[51], pour le problème de la petite phylogénie, en prenant comme données d'entrée un ensemble de génomes basé sur l'ordre des gènes et une structure arborescente (Références: [78])

Desenvolvimento de uma abordagem baseada em busca em vizinhança variável para o problema da grande filogenia tomando como dados de entrada um conjunto de genomas baseados na ordem dos genes (Referências: [78]). ||| •

Développement d'une approche de recherche de quartier variable pour le problème de la phylogénie grande en utilisant comme données d'entrée un ensemble de génomes basés sur l'ordre des gènes (Références: [78])

Realizaram-se experimentos usando datasets importantes da literatura e se compararam os resultados com outras abordagens da literatura para o problema da grande e pequena filogenia. ||| •

Les expériences ont été effectuées en utilisant des ensembles de données importants de la littérature et les résultats ont été comparés avec d'autres approches dans la littérature pour le problème de la phylogénie grande et petite

No Capítulo 4, são apresentados os novos algoritmos evolutivos propostos para o problema do cálculo da distância de reversão, também são apresentados os resultados dos experimentos, bem como uma comparação estatística destes resultados usando os testes de Friedman e Holm. |||

Au chapitre 4, les nouveaux algorithmes d'évolution proposés pour le problème du calcul de la distance de réversion sont présentés, les résultats des expériences sont présentés ainsi qu'une comparaison statistique de ces résultats en utilisant les tests Friedman et Holm



Universidade de Brasília
Biblioteca Central

No Capítulo 5, são apresentados as abordagens heurísticas para lidar com o problema da pequena e grande filogenia, também são apresentados resultados dos experimentos que consistem em novas árvores filogenéticas para os datasets Campanulaceae e Hemiascomycetes. |||

Assim, a ordem dos genes de um organismo pode ser interpretada, em notação de string, como uma permutação sem sinal $\pi = \pi_1, \pi_2, \dots$ |||

Seja $\pi = 5, 3, 2, 6, 4, 1$ uma permutação sem sinal de comprimento 6, onde esta sequência de números representa a ordem dos genes de um organismo. |||

De acordo com esta definição, os elementos de qualquer permutação π no intervalo de posições $[i, j]$ são revertidos dentro deste intervalo como ação de uma reversão, escrito em notação funcional como $\rho_{i..j} \circ \pi$, onde o símbolo \circ denota a composição de funções. |||

A distância de reversão entre duas permutações sem sinal π e σ é o mínimo número de reversões para transformar π em σ , e o problema de encontrar esta distância é conhecido como Problema da Distância de Reversão (PDR). |||

Uma vez que este problema é equivalente a transformar $\sigma^{-1} \circ \pi$ em 1 , podemos expressar o PDR como o problema de encontrar a distância de

Au chapitre 5, les approches heuristiques pour traiter le problème de la phylogénie petite et grande sont présentées, les résultats des expériences consistant en de nouveaux arbres phylogénétiques pour les jeux de données Campanulaceae et Hemiascomycètes sont également présentés

Ainsi, l'ordre des gènes d'un organisme peut être interprété, en notation de chaîne, comme une permutation non signée $\pi = \pi_1, \pi_2, \dots$

Soit $\pi = 5, 3, 2, 6, 4, 1$ une permutation non signée de longueur 6, où cette séquence de nombres représente l'ordre des gènes d'un organisme

Selon cette définition, les éléments de toute permutation π dans la plage de positions $[i, j]$ sont inversés dans cet intervalle comme une action de réversion, écrit en notation fonctionnelle comme $\rho_{i..j} \circ \pi$, où le symbole \circ désigne Composition des fonctions

La distance d'inversion entre deux permutations non signées π et σ est le nombre minimal d'inversions pour transformer π en σ , et le problème de trouver cette distance est connu comme le problème de la distance d'inversion (PDR)

reversão entre uma permutação σ e 1 , este problema é conhecido como Ordenação de Permutações sem Sinal por Reversões (OPSSR). |||



Universidade de Brasília
Biblioteca Central

Étant donné que ce problème équivaut à transformer $\sigma^{-1} \circ \pi$ en 1 , nous pouvons exprimer la PDR comme le problème de trouver la distance de

Seja ρ uma reversão que transforma π em π_0 , e deixe $b(\pi)$ denotar o número de pontos de quebra de uma permutação sem sinal π . |||

Um grafo de pontos de quebra (grafo de ciclos) $G(\pi)$ da permutação π é um grafo de arestas coloridas derivado das adjacências e pontos de quebra de π o qual tem $n + 2$ vértices, um vértice para cada elemento de π incluindo os pivôs. |||

Podemos verificar facilmente que a única permutação sem nenhuma aresta é a permutação identidade, uma vez que não tem pontos de quebra e todos seus vértices consecutivos são adjacentes. ||| Nous pouvons

De fato, para qualquer ponto de quebra formado por um vértice π_i existe uma aresta preta, e uma aresta cinza que vai até um vértice não consecutivo π_j tal que $\pi_j \sim \pi_i$. |||

Pode-se verificar que para um grafo $G(\pi)$, o qual foi gerado a partir de uma permutação sem sinal, existem muitas decomposições em ciclos diferentes. |||

réversion entre une permutation σ et 1 , ce problème est connu sous le nom de Permutations de Commande sans Signaux par Reverses (OPSSR)

Soit ρ un renversement qui transforme π en π_0 et que $b(\pi)$ indique le nombre de points d'arrêt d'une permutation non signée π

Un graphique des points d'arrêt (graphique des cycles) $G(\pi)$ de la permutation π est un graphique des arêtes colorées dérivées des adjacences et des points d'arrêt de π qui a $n + 2$ sommets, un sommet pour chaque élément de π y compris le Pivots

facilement vérifier que la seule permutation sans bord est la permutation d'identité, car elle n'a pas de points d'arrêt et tous ses sommets consécutifs sont adjacents

En fait, pour tout point de rupture formé par un sommet π_i , il y a un bord noir et un bord gris qui remonte à un sommet non consécutif π_j tel que $\pi_j \sim \pi_i$

On peut vérifier que pour un graphique $G(\pi)$, qui a été généré à partir d'une permutation non signée, il existe de nombreuses décompositions dans différents cycles



**Universidade de Brasília
Biblioteca Central**

Assim, para cada permutação sem sinal π podemos construir $2n$ permutações com sinal diferentes, isto é feito atribuindo um sinal positivo ou negativo a cada elemento de π . |||

.-----
Neste caso o problema de determinar a distância de reversão entre uma permutação $\rightarrow\pi$ e a permutação identidade é conhecido como Ordenação de Permutações com Sinal por Reversões (OPCSR). ||| Dans ce cas,

.-----
Esta transformação leva a permutações cujos grafos de pontos de quebra são de tal forma que cada vértice tem no máximo grau dois, quer dizer, exatamente uma aresta preta e uma aresta cinza (ver Figura. |||

.-----
Hannenhalli e Pevzner [45] propuseram uma relação simples ($d(\rightarrow\pi) = b(\rightarrow\pi) - c(\rightarrow\pi) + h(\rightarrow\pi) + f(\rightarrow\pi)$) para calcular de forma exata a distância de reversão de permutações com sinal, onde $h(\rightarrow\pi)$ e $f(\rightarrow\pi) \in \{0, 1\}$ são noções que indicam se uma permutação é difícil de ser ordenada. ||| Hannenhalli et Pevzner [45] ont

.-----

Ainsi, pour chaque π de permutation sans signe peut construire des permutations avec $2n$ signal différent, cela se fait en attribuant un signe positif ou négatif à chaque élément de π

le problème de la détermination de la distance entre la permutation d'inversion $\rightarrow\pi$ et la permutation d'identité est connue sous le nom Trier Permutations avec inversion de signe (OPCSR)

Cette transformation conduit à des permutations dont les graphes des points de rupture sont tels que chaque sommet a au plus deux de qualité, qui est, exactement un bord noir et un bord de gris (voir Fig

proposé une relation simple ($d(\rightarrow\pi) = b(\rightarrow\pi) - c(\rightarrow\pi) + h(\rightarrow\pi) + f(\rightarrow\pi)$) afin de calculer la distance exacte avec permutation d'inversion de signe, où $h(\rightarrow\pi)$ $f(\rightarrow\pi) \in \{0, 1\}$ sont des notions qui indiquent si une permutation est difficile d'être ordonnée



Universidade de Brasília
Biblioteca Central

Esta relação levou ao desenvolvimento de um algoritmo de tempo quadrático ($O(n^2)$) para calcular a distância de reversão de permutações com sinal, e que tem complexidade $O(n^4)$ quando adicionalmente tem que ser calculada a sequência de reversões para ordenar a permutação com sinal. Logo, Bergeron [10] propôs uma apresentação elementar da teoria de Hannenhalli e Pevzer [45] que age diretamente sobre a permutação com sinal a ser ordenada. |||

.-----

A ideia para ordenar uma permutação sem sinal $\rightarrow \pi$ usando o número mínimo de reversões é a seguinte: (1) aplicar reversões sobre "pares ordenados", logo no final teremos como resultado uma permutação com todos os elementos positivos; (2) se no passo anterior a permutação não está ordenada, então aplicar reversões sobre "obstáculos" de forma que novos "pares ordenados" sejam criados. |||

.-----

Logo, o escore da reversão $\rho_{3..5}$ é 4 porque depois de ser aplicada sobre $\rightarrow \pi$ temos a seguinte permutação $\rho_{3..5} \circ \rightarrow \pi = 0, 4, 2, 3, -1, -5, 6$ com os seguintes pares ordenados $(0, -1)$, $(2, -1)$, $(4, -5)$, e $(-5, 6)$. |||

.-----

Cette relation a conduit à l'élaboration d'un algorithme quadratique ($O(n^2)$) pour calculer la distance de l'inversion de permutations du signal, qui a une complexité en $O(n^4)$ lorsque doivent en outre être séquence estimée de reprises pour trier les permutations Avec signe. Par conséquent, Bergeron [10] a proposé une présentation élémentaire de la théorie Hannenhalli Pevzer et [45] qui agit directement sur le signal de permutation à trier

L'idée de commander une permutation non signée $\rightarrow \pi$ en utilisant le nombre minimum de reprises est la suivante: (1) appliquer les reprises sur les « paires ordonnées », puis à la fin, nous aurons à la suite d'une permutation avec tous les éléments positifs; (2) si, à l'étape précédente, la permutation n'est pas commandée, puis appliquez des renversements sur les "obstacles" afin de créer de nouvelles "paires ordonnées"

Ensuite, le score de réversion $\rho_{3..5}$ est 4 car, après avoir été appliqué sur $\rightarrow \pi$, nous avons la permutation suivante $\rho_{3..5} \circ \rightarrow \pi = 0, 4, 2, 3, -1, -5, 6$ avec Les paires commandées suivantes $(0, -1)$, $(2, -1)$, $(4, -5)$ et $(-5, 6)$



Universidade de Brasília
Biblioteca Central

Esta estratégia dá como resultado uma permutação com todos seus elementos positivos. |||

.-----

Em caso de que a permutação não fique ordenada temos que aplicar outra estratégia para lidar com esse problema, que será explicada a continuação. |||

.-----

Algoritmo 1: Estratégia Básica para Ordenar Permutações com Sinal
Entrada: Uma permutação com sinal $\rightarrow \pi$
Saída: Uma permutação com sinal $\rightarrow \pi$ que só tem elementos positivos 1

.-----

Aplicar a reversão ρ sobre $\rightarrow \pi$;
Ordenação de Permutações com Sinal que tem só Elementos Positivos Seja $\rightarrow \pi = 0, \pi_1, \pi_2, .$ |||

„-----

Cette stratégie entraîne une permutation avec tous ses éléments positifs

Dans le cas où la permutation n'est pas ordonnée, nous devons appliquer une autre stratégie pour résoudre ce problème, ce qui sera expliqué dans la suite

enquanto $\rightarrow \pi$ tenha pares orientados faça 2 |||

Une permutation avec le signal $\rightarrow \pi$
Sortie: Une permutation avec le signal $\rightarrow \pi$ qui n'a que des éléments positifs 1
alors que $\rightarrow \pi$ a paires orientées 2

Appliquer la réversion ρ sur $\rightarrow \pi$;
Commande des permutations avec le signal qui n'a que des éléments positifs
Soit $\rightarrow \pi = 0, \pi_1, \pi_2$

ANEXO J – LICENÇAS DE USO DE SOFTWARE

PDF MINER (usado no Conversor)

Copyright (c) 2004-2016 Yusuke Shinyama <yusuke at shinyama dot jp>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

GENSIM (usado no Sumarizador)

aRe-Technologies/gensim is licensed under the

GNU Lesser General Public License v2.1

Primarily used for software libraries, the GNU LGPL requires that derived works be licensed under the same license, but works that only link to it do not fall under this restriction. There are two commonly used versions of the GNU LGPL.

TEXTBLOB (Usado no Tradutor)

sloria/TextBlob is licensed under the

MIT License

A short and simple permissive license with conditions only requiring preservation of copyright and license notices. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

Copyright 2013-2017 Steven Loria