

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 22 日現在

機関番号：12601

研究種目：基盤研究（A）

研究期間：2009～2012

課題番号：21240021

研究課題名（和文） 包括的な翻訳情報資源を実現する統合翻訳支援サイトの構築

研究課題名（英文） Developing an integrated translation-aid site that provides comprehensive reference resources for translators

研究代表者

影浦 峡（KAGEURA KYO）

東京大学・大学院教育学研究科・教授

研究者番号：00211152

研究成果の概要（和文）：(1) レファレンス・ツールにおける「包括性」の概念および包括性を実現するための要件を明らかにした。(2) 専門語彙クローラーと対訳・関連多言語アーカイブクローラーを開発し、機能的包括性を有するレファレンス情報資源を構築した。(3) 翻訳情報資源を提供する統合翻訳支援サイト「みんなの翻訳」(<http://trans-aid.jp/>)を開発・公開し、一般利用に提供し翻訳情報資源の有効性を検証した。

研究成果の概要（英文）：This research (1) clarified the concept of “comprehensiveness” for reference tools and concrete factors that should be incorporated into reference tools; (2) developed a comprehensive bilingual terminology crawler and a parallel and comparable archive constructor and constructed reference resources for translators; and (3) developed and made publicly available the integrated translation-aid environment Minna no Hon'yaku (<http://trans-aid.jp/>).

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	9,200,000	2,760,000	11,960,000
2010年度	9,400,000	2,820,000	12,220,000
2011年度	9,900,000	2,970,000	12,870,000
2012年度	6,100,000	1,830,000	7,930,000
年度			
総計	34,600,000	10,380,000	44,980,000

研究分野：総合領域

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：翻訳支援・翻訳情報資源・レファレンス・アーカイブ・多言語処理

1. 研究開始当初の背景

オンライン文書の翻訳発信の展開が進む中、様々な翻訳支援ツールや翻訳＝ニーズマッチングのサービスが提供されてきたが、レファレンス・コンテンツを体系的に提供するサービスはほとんどなかった。また、レファレンス・コンテンツに求められる具体的な要件についても明らかにされていなかった。我々はすでにオンライン翻訳者支援の枠組

みを構築していたが、それを、この点から展開することは実利用を促進する観点からの急務でもあった。

2. 研究の目的

大きく以下の三点からなる。

(1) 翻訳支援と翻訳情報発信を行う統合オンラインサイトの構築。

(2) 翻訳情報資源を階層的に編成・提供する基盤となる理論の構築と技術の開発。

(3) 包括的なレファレンス資源に求められる具体的な包括性要件の理論的整理と、それに従ったレファレンス情報資源の構築。

3. 研究の方法

階層的な翻訳情報資源の編成および包括的なレファレンス資源構築の基盤となる理論的研究においては、既往のレファレンスツール論・辞書学の知見を参照するとともに、翻訳者への聞き取り・翻訳行動の観察を行う。

翻訳支援・翻訳情報発信統合オンラインサイトの構築、翻訳情報資源の階層的構成メカニズムの開発、包括的なレファレンス情報資源の構築に関しては、理論的検討の成果を踏まえつつ、実システムを開発するという工学的アプローチを採用する。

評価については、要素技術の評価は条件を整えて量的な観点から実験的に行うが、包括性および統合サイトの評価は実利用を通して質的な観点からも行う。

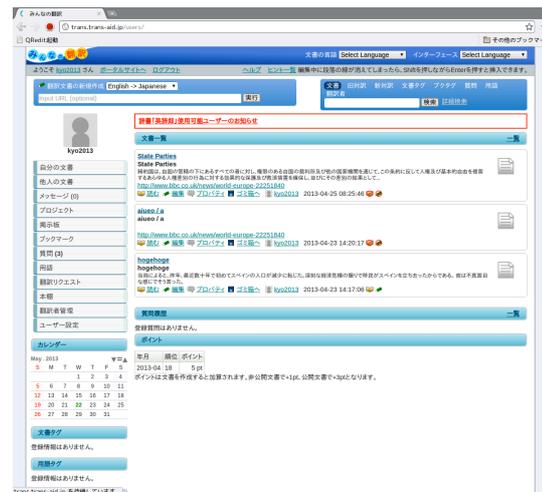
4. 研究成果

(1) オンライン翻訳支援・情報発信プラットフォーム「みんなの翻訳」(<http://trans-aid.jp/>)を構築・公開した。現在、英日・日英のほか、中英・英中、日中・日英、韓英・英韓、韓日・日韓、日独・カタロニア英をカバーしており、インタフェースもこれら言語に基本的に対応している。当初目標である定常ユーザ 50 人を達成している(登録者は 2013 年 5 月現在、2500 人強)。翻訳文書量は日英対訳単語数で 110 万を超えている。

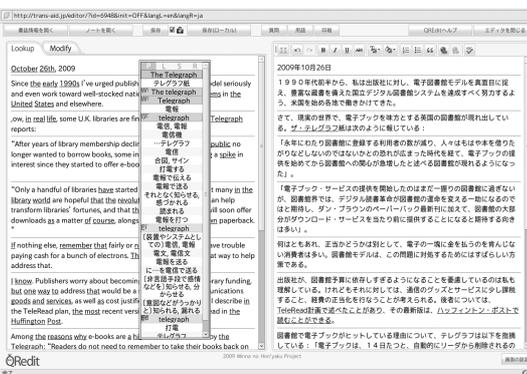
みんなの翻訳ポータル



ユーザスペース



翻訳支援エディタ QRedit



利用者は日本のほか、ドイツ、英国、スペイン(カタロニア)、韓国、中国にわたり、作品社や岩波書店の翻訳書出版にも一部利用されている。また、神戸女学院大学翻訳コース、東京外国語大学翻訳コース、東京大学の翻訳ゼミでも活用されており、ボランティア・オンライン翻訳の一つの形を示すものとなっている。

(2) オンライン翻訳者のレファレンス情報資源参照行動をもとに、基本辞書引き、詳細辞書引き、Google 検索、Google 辞書検索、既訳関連文書検索(翻訳メモリ検索)、ユーザ用語集検索を区別し、翻訳支援エディタからこれらの情報に、区別を翻訳者側で意識しコントロールしながらシームレスに検索する機能とインタフェースを構築した。

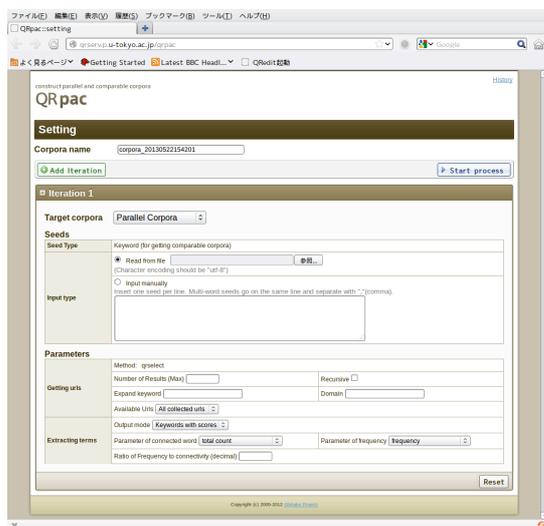
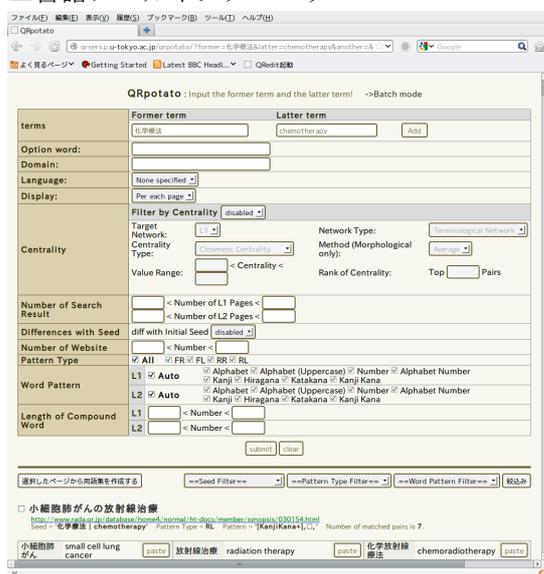
(3) レファレンス・ツールにおける「包括性」の概念を、(a) 絶対的包括性と相対的包括性、(b) 理念的包括性、社会的包括性、機能的包括性の観点から整理し、オンライン翻訳支援環境においては、(a)においては相対的包括性

が、(b)においては社会的包括性と機能的包括性が重要であることを示すとともに、包括性を実現するための要件を明らかにした。

また、対訳専門語彙、対訳文書について、Web 上から情報を包括的にクロールする手法を開発した。対訳専門語彙については、社会的包括性をターゲットとし、日英対訳で、シード辞書約 20 万語を用い、延べ 600 万語、異なり 300 万語、精度約 80% で用語辞書資源を収集した。不特定多数への実利用への有効性について現在、検証を行っている。対訳文書については、個人ベースでの機能的包括性をターゲットとし、翻訳者個人が自らの翻訳プロファイルに応じた関連規訳文書を収集するためのクローラを公開し、試験利用に供している。

用語対訳クローラ

二言語アーカイブクローラ



5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 7 件)

1. Kyo Kageura and Ryo Murayama (2013) “Web-based archiving of parallel and comparable documents,” *Studies in Lifelong Learning Infrastructure Management*, 37, p. 47-56. (査読有り)
2. Kyo Kageura, Takeshi Abekawa, Masao Utiyama, Miori Sagara and Eiichiro Sumita (2011) “Has translation gone online and collaborative? An experience from Minna no Hon’yaku,” *Theme in Translation Studies*, 10, p. 47-72. (招待論文・査読無し)
3. Kyo Kageura (2010) “Analysing the status of borrowed morphemes in terminological structure,” *Terminology*, 16(2), p. 181-216. (査読有り)
4. Emmanuel Morin, Beatrice Daille, Koichi Takeuchi and Kyo Kageura (2010) “Brains, not brown: the use of ‘smart’ comparable corpora in bilingual terminology mining,” *ACM Transactions on Speech and Language Processing*, 7(1), paper 1. (査読有り)
5. 影浦峯, 阿辺川武, 内山将夫 (2010) 「みんなの翻訳は何を目指すか」 *AAMT Journal*, 48, p. 8-10. (査読無し)
6. Kyo Kageura (2009) “Computing the potential lexical productivity of head elements in nominal compounds using the textual corpus,” *Progress in Informatics*, 5, p. 49-56. (査読有り)
7. 内山将夫, 阿辺川武, 隅田英一郎, 影浦峯 (2009) 「みんなの翻訳-ボランティア翻訳者の支援と翻訳の共有のための web サイト」 *AAMT Journal*, 45, p. 11-14. (査読無し)

[学会発表] (計 26 件)

1. Martin Thomas, Bogdan Babich, Anthony Hartley, Masao Utiyama and Kyo Kageura (2012) “MNH-TT: a collaborative platform for translator training,” *Translating and the Computer* 34, November 29-30, UK.
2. Kyo Kageura and Ryo Murayama (2012) “QRpac: user-driven archiving of parallel and comparable corpora from the web,” *The 14th International Conference on Asian Digital Libraries*, November 12-15, Taiwan.
3. Bogdan Babich, Anthony Hartley, Kyo Kageura, Martin Thomas and Masao Utiyama (2012) “Scaffolding, capturing, preserving interactions in collaborative translation,” *2012 LTTC International Conference*, April 28-29, Taiwan.
4. Takuma Asaishi and Kyo Kageura (2011) “Comparative analysis of the motivatedness

structure of Japanese and English terminologies,” *The 9th International Conference on Terminology and Artificial Intelligence*, November 8-10, France.

5. Takeshi Abekawa and Kyo Kageura (2011) “Using seed terms for crawling bilingual terminology lists on the Web,” *Translating and the Computer*, November 17-18, UK.

6. Kyo Kageura and Takeshi Abekawa (2011) “On the concept of ‘comprehensiveness’ in information services: the case of the online translation aid and hosting service Minna no Hon’yaku,” *Asia Pacific Conference on Library and Information Education and Practice*, June 22-24, Malaysia.

7. Kyo Kageura (2011) “On an infrastructure for online coordinated translation training: extending the online translation aid system Minna no Hon’yaku for training translators,” *The 1st International Conference on Law, Translation and Culture*, August 28-30, China. (招待講演)

8. Masao Utiyama, Takeshi Abekawa, Kyo Kageura and Eiichiro Sumita (2011) “A strategic supply of reference resources and lookup environments in online translation hosting site,” *Asialex 2011*, August 22-24, Japan.

9. Kyo Kageura (2011) “The emerging landscape of translation and the case for an open online translation environment,” *The 1st International Conference on Law, Language and Discourse*, August 20-21, Hong Kong, China. (招待講演)

10. 内山将夫, 阿辺川武, 影浦峽 (2011) 「みんなの翻訳第三報」言語処理学会第17回年次大会, 3月8-10日.

11. 影浦峽, 阿辺川武 (2011) 「翻訳者向けレファレンス・ツールにおける『包括性』概念をめぐって」言語処理学会第17回年次大会, 3月8-10日.

12. 影浦峽 (2010) 「多言語情報流通環境と『翻訳』の場」関西トランスレーションスタディーズ研究会, 10月2日.

13. Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura (2010) “Helping volunteer translators, fostering language resources,” *COLING 2010 Workshop on People’s Web Meets NLP*, August 28, China.

14. Kyo Kageura (2010) “Being theoretical is being practical: multiword units and terminological structure revitalised,” *COLING 2010 Workshop on Multiword Expressions*, August 28, China. (招待講演)

15. Mohammad Daoud, Kyo Kageura, Christian Boitet, Asanobu Kitamoto and Daoud Daoud (2010) “Passive and active contribution to multilingual lexical resources through online cultural activities,” *The 6th IEEE International Conference on Natural Language Processing and*

Knowledge Engineering, August 21-23, China.

16. Mohammad Daoud, Kyo Kageura, Christian Boitet, Asanobu Kitamoto and Mathieu Mangeot (2010) “Multilingual lexical network from the archive of Digital Silk Road,” *OntoLex 2010*, August 21-22, China.

17. Takeshi Abekawa, Masao Utiyama, Eiichiro Sumita and Kyo Kageura (2010) “Management and use of terminological resources for distributed users in the translation hosting site Minna no Hon’yaku,” *The 14th Euralex Congress*, July 6-10, The Netherlands.

18. Kyo Kageura, Masao Utiyama, Takeshi Abekawa and Eiichiro Sumita (2010) “Minna no Hon’yaku: hosting, archiving and promoting (human and machine) translations,” *TAUS Tokyo Executive Forum*, April 14-15, Japan.

19. 影浦峽, 阿辺川武, 内山将夫, 隅田英一郎 (2010) 「オンライン協調翻訳環境におけるユーザ用語管理メカニズム」言語処理学会第16回年次大会, 3月8-11日.

20. 内山将夫, 阿辺川武, 影浦峽, 隅田英一郎 (2010) 「みんなの翻訳第二報」言語処理学会第16回年次大会, 3月8-11日.

21. Takeshi Abekawa and Kyo Kageura (2009) “QRpotato: A system that exhaustively collects bilingual technical term pairs from the web,” *3rd International Universal Communication Symposium*, December 3-4, Japan.

22. Mohammad Daoud, Christian Boitet, Kyo Kageura, Asanobu Kitamoto, Mathieu Mangeot and Daoud Daoud (2009) “Constructing multilingual preterminological graphs using various online-community resources,” *8th International Symposium on Natural Language Processing*, October 21-22, Thailand.

23. Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura (2009) “Hosting volunteer translators,” *MT Summit XII*, August 27-29, Canada.

24. Emmanuel Prochasson, Emmanuel Morin and Kyo Kageura (2009) “Anchor points for bilingual lexicon extraction from small comparable corpora,” *MT Summit XII*, August 27-29, Canada.

25. Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura (2009) “Minna no Hon’yaku: a website for hosting, archiving and promoting translations,” *Translating and the Computer 31*, November 19-20, UK.

26. Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura (2009) “Features and characteristics of Minna no Hon’yaku,” *Open Translation Tools*, June 22-24, The Netherlands.

[図書] (計8件)

1. Kyo Kageura and Takeshi Abekawa (to appear) “The place of comparable corpora in

providing terminological reference information to online translators: a strategic framework,” Serge Sharoff, Pierre Zweigenbaum and Reinhard Rapp (eds.) *Building and Using Comparable Corpora*. Berlin: Springer.

2. Kyo Kageura (2012) *The Quantitative Analysis of the Dynamics and Structure of Terminologies*.

3. 影浦峽 (2011) 「情報学と語彙」 齊藤倫明・石井正彦編『これからの語彙論』ひつじ書房, p. 201-212.

4. 影浦峽 (2010) 「用語抽出」『言語処理学事典』共立出版, p. 350-353.

5. 影浦峽 (2010) 「ターミノロジー」『言語処理学事典』共立出版, p. 94-95.

6. 影浦峽 (2010) 「語彙量の推定」『計量国語学事典』朝倉書店, p. 118-122.

7. 影浦峽 (2010) 「使用率の分布」『計量国語学事典』朝倉書店, p. 112-115.

8. 影浦峽 (2009) 「かたちから言葉を見る」新井紀子『数学は言葉』東京図書, p. 155-172.

〔産業財産権〕

○出願状況 (計0件)

○取得状況 (計0件)

〔その他〕

受賞

Asian Association for Machine Translation Nagao Award (2010) July 14.

雑誌記事

影浦峽 (2012) 「巻頭言 翻訳を通して『言葉』を見ると」『自然言語処理』19(2), p. 63-64.

ホームページ等

<http://trans-aid.jp/>

<http://qrserv.p.u-tokyo.ac.jp/qrpotato/>

<http://qrserv.p.u-tokyo.ac.jp/qrpac/>

(後2つは限定公開)

6. 研究組織

(1) 研究代表者

影浦峽 (KAGEURA KYO)

東京大学・大学院教育学研究科・教授

研究者番号：00211152

(2) 研究分担者

阿辺川武 (ABEKAWA TAKESHI)

情報・システム研究機構・特任准教授

研究者番号：00431776

内山将夫 (UTIYAMA MASAO)

情報通信研究機構・言語翻訳グループ

・主任研究員

研究者番号：70293496

(3) 連携研究者

佐藤理史 (SATO SATOSHI)

名古屋大学・大学院工学系研究科・教授

研究者番号：30205918

宇津呂武仁 (UTSURO TAKEHITO)

筑波大学・大学院システム情報工学研究科・准教授

研究者番号：90263433

竹内孔一 (TAKEUCHI KOICHI)

岡山大学・大学院自然科学研究科・講師

研究者番号：80311174

相澤彰子 (AIZAWA AKIKO)

国立情報学研究所・教授

研究者番号：90222447

戸田慎一 (TODA SHINICHI)

東洋大学・社会学部・教授

研究者番号：10183493