様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 １６ 日現在

機関番号：３２６８９
研究種目：基盤研究(C)
研究期間：2011～2014
課題番号：２３５０１１１５

研究課題名（和文）大規模な注釈つきコーパス分析のための直感的コーパスツール開発

研究課題名（英文）Development of an Intuitive Corpus Tool for Analysis of Large-Scale, Fully-Annotated Corpora

研究代表者

Ａｎｔｈｏｎｙ　Ｌａｕｒｅｎｃｅ（Anthony, Laurence）

早稲田大学・理工学術院・教授

研究者番号：１０２５８２０４

交付決定額（研究期間全体）：（直接経費）　　3,800,000円

研究成果の概要（和文）：この研究の最終目的はコーパス分析ソフト「AntConc」に新たな機能を加え、大規模な注釈つきコーパスをより簡単に分析できることである。よって、コーパス利用者の多くのニーズに対応することができる。この目的を果たすため、３つの目標を設定した。目標１：「AntConc」をPython言語で書き直し、より簡単に機能を加えることができ、操作性を高める。目標２：「AntConc」の利用するデータベースを再設計し、１０億語のコーパスを市販のノートパソコンで処理できるようにする。目標３：新しいタガーツールを作成し、短文レベルや文書レベルでのテキスト処理を可能にする。研究結果として各目標をほぼ確実に実現した。

研究成果の概要（英文）：The aim of this project was to redesign and add new functionality to the AntConc corpus toolkit so that it could respond to the increasing demands of corpus linguists for sophisticated tools that can handle very large, fully-annotated corpora in an easy and intuitive way. To this aim, all three core goals of the project were completed: 1) AntConc has been completely re-written in the Python programming language making it significantly easier to add new functionality and a much improved user experience; 2) AntConc has been redesigned with a novel backend database to allow it to process massive annotated corpora of over 1 billion words on a standard laptop computer; 3) New modules have been created to allow users to easily process texts at the sentence and/or discourse level.

研究分野：教育工学

キーワード：AntConc　コーパス　コンコーダンス　教育工学　語彙　corpus　concordance　educational technology

1.研究開始当初の背景
(1) コーパス言語分析の結果によって、言語理論の発展、新言語指導法・学習法、便利な辞書・教科書そして効率の良い・正確な翻訳方法ができた。

(2) 一方、近代のコーパス研究ではコーパスそのものが益々大きくなり、注釈つきのコーパスが主流となった。この変化により、２つの大きな課題が現れた。１つ目はソフトウェアの限界である。現在、パソコンで使えるソフトウェアは１００万語以下のコーパスに対応できているが、これ以上のコーパスになると分析処理が遅くなり、ソフトウェアそのものがクラッシュすることがある。また、多くのパソコンで使えるソフトウェアは注釈つきのコーパスに対応していないので、このようなコーパスを分析したい場合、ブラウザーを通して公開されているもののみ分析できる。
２つ目の問題はコーパスデータに注釈を付けるソフトウェアの使用法と操作性である。現在、多くのタガー（注釈を付けるソフトウェア）が存在しているが、実際に使えるまで、関連ソフトウェアのダウンロード、辞書設定、コマンドラインでの起動指令などの複雑な手順に従うことになる。よって、研究者以外のコーパスに興味のある者（教員・学習者など）はほとんど触れていない。その上、世界一の正確さのあるタガーをブラウザー上で使うことが多いのでファイル数の多いコーパスに品詞タグまたは文書レベルの注釈を付けたい場合、ほぼ不可能になる。

(3) 上で説明したとおり、多くのコーパスツールはコーパス学者（研究者）をターゲットにしているが、コーパスの普及により、コンピューターに慣れている教員・学生・言語学者以外の者にも使える分かり易いコーパス分析ソフトウェアが必要となる。

２．研究の目的
この研究では３つの目標を設定した。
(1) 過去に開発した世界基準となるコーパス分析ソフトウェア「AntConc」を Python 言語で書き直す。よって、Python のオブジェクト適応性と構文により、新機能をより簡単に加えることができる。また、Python に対応している高度なインターフェース設計環境が存在しているので、研究者・教員・学習者などのコンピューターに自信がないユーザーに対応できる分かり易いインターフェースが作成できる。
(2)「AntConc」のバックエンドデータベース（内蔵されているデータベース）を再設計し、一般人の持つノートパソコンで１０億語以上の大規模のコーパスを分析できるようなソフトウェアにする。
(3)「AntConc」と同様なワンクリック起動出来るタガー（注釈を付ける新ソフトウェア）を開発する。よって、多くの研究者・教員・学習者などのコーパス利用者が小・中・大規模のコーパスに品詞タグや文書レベルでの注釈を付けることができる。

３．研究の方法
(1) 新バージョンの「AntConc」の開発環境の以下の通りでした：

言語：Python 2.7.6
インターフェース設計環境：PyQt
プログラミング環境：Eclipse (Juno) + PyDev
コンパイラー・パケジャー：PyInstaller

Python と PyQt での開発により、新バージョンの「AntConc」が以前のバージョンとどうように Windows、 Macintosh OS X そして Linux コンピューターで使うことができた。

(2) イギリスのランカスター大学の研究者の協力を得て、多くの実験・テスティングを元に「AntConc」のバックエンドデータベースとして、HDF5 技術と Sqlite 技術を使い、融合したデータベース構築（通称 AntHDF5）をした。HDF5 と Sqlite が両方ポータブル （持ち運び）のものであるので、AntConc」も以前のバージョンと同様にポータブルになる。

(3) イギリスのランカスター大学の研究者の協力を得て、上記で説明した「AntConc」開発環境で複数の言語対象のポータブルタガーを同時に開発した。

３．研究成果
(1) 新バージョンの「AntConc」：図１は開発された「AntConc」のファイル表示ツールのスクリーンショットを示す。開発された「AntConc」には簡単に開発できるモジュールが含まれ、以前のバージョンのツールバーに表示される。よって、ソフトウェアの拡張性が増してくる。開発された「AntConc」は Windows、 Macintosh OS X そして Linux コンピューターで「ネイティブアプリ」として使える。

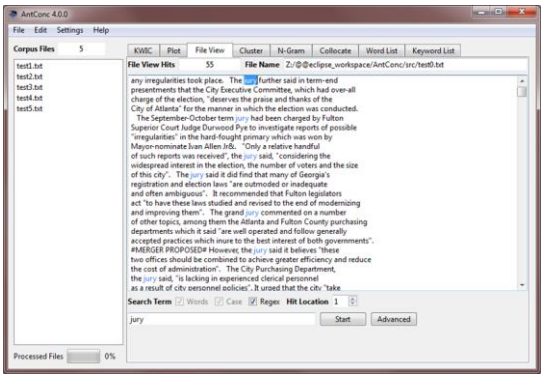

図１　「AntConc」のスクリーンショット

(2) 「AntConc」のバックエンドデータベース：新バックエンドデータベースの効果を調

べるため、いくつのテストを行った。表１には以前のバージョンの「AntConc」と新バージョンの「AntConc」のスピードを比較するテスト結果が示されている。このテストでは Brown Corpus と British National Corpus (BNC) での"the"（コーパスの一番頻度の高い単語）を各バージョンで検索し、コンコーダンス結果を表示するまでの時間を図った。表１で示すように AntHDF5 のデータベース構築により、以前のバージョンの「AntConc」と比べ、新バージョンでは千倍以上の検索スピードを得ることができた。

表１　「AntConc」の検索スピードの比較

バージョン	テストコーパス	"the"の検索時間
AntConc 3.4.3	Brown (1 m)	13.11 sec
AntConc 3.4.3	BNC (100 m)	1210.13 sec
AntConc 4.0	Brown (1 m)	0.07 sec
AntConc 4.0	BNC (100 m)	2.94 sec

(m = １万語)

(3) タガー（注釈つきツール）：図２は開発された「TagAnt」のスクリーンショットを示す。「TagAnt」は英語以外にフランス語、ポルトガル語、ドイツ語などの言語に対応している。

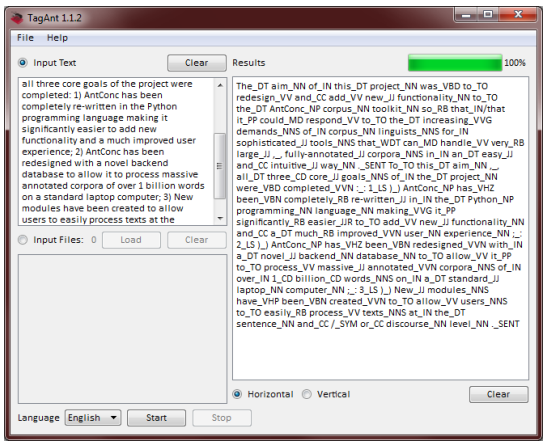

図２　「TagAnt」のスクリーンショット

図３は開発された「ClawsAnt」のスクリーンショットを示す。「ClawsAnt」はイギリスのランカスター大学で開発された英語対応の CLAWS タガーのポータブル化されたものである。CLAWS は世界一の正確さを持つタガーと知られている。

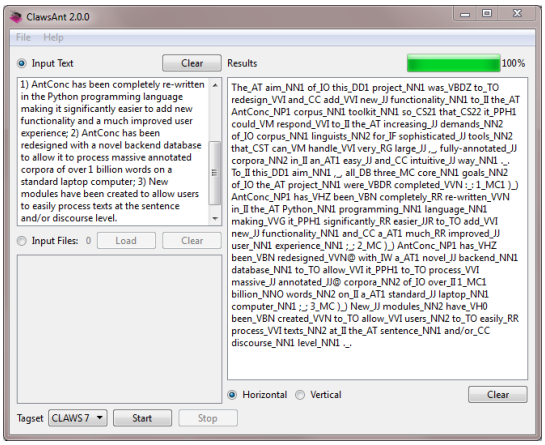

図３　「ClawsAnt」のスクリーンショット

図４は開発された「SegmentAnt」のスクリーンショットを示す。「SegmentAnt」は日本語・中国語などのアジア言語に対応している単語分け・タガーツールである。

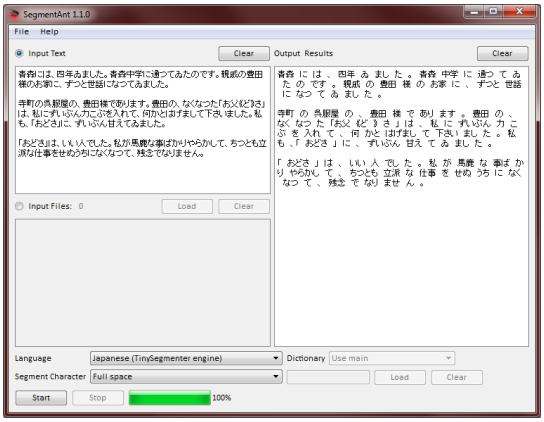

図４　「SegmentAnt」のスクリーンショット

今回の開発したツールはすべて以下のサイトで公開されている。
http://www.laurenceanthony.net/software.html

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　24　件）

1. Anthony, L. and Baker, P.: "ProtAnt: A tool for analysing the prototypicality of texts" International Journal of Corpus Linguistics (in press), 査読有
2. Vidler, J., Rayson, P., Anthony, L. Scott A., and Mariani, J.: "Dealing With Big Data Outside Of The Cloud: GPU Accelerated Sort" Proceedings of the Language Resources and Evaluation Conference (LREC 2014), May 26-31, 2014. Harpa Conference Centre, Reykjavik, Iceland. 14-16 (2014), 査読有
3. Cheng, A. and Anthony, L.: "ESP research in Asia" English for Specific Purposes 33. 1-3 (2014), 査読有
4. Chujo, K., Anthony, L., Oghigian, K., and Yokota, K.: "Teaching remedial grammar through

Data-Driven Learning using AntPConc" Taiwan International ESP Journal 5:2. 65-90 (2014), 査読有

5. Anthony, L. and Bowen, M.: "The Language of Mathematics: A Corpus-based Analysis of Research Article Writing in a Neglected Field" Asian ESP Journal 9:2. 5-25 (2013), 査読有
6. Anthony, L.: "A critical look at software tools in corpus linguistics" Linguistic Research 30:2. 141-161 (2013), 査読有
7. Chujo, K., Anthony, L., Utiyama M., and Nishigaki, C.: "WebParaNews を利用した Web 版 DDL 教材の開発" 日本大学生産工学部研究報告Ｂ 6:46. 27-37 (2013), 査読無
8. Nation, P. and Anthony, L.: "Mid-frequency readers" Journal of Extensive Reading 1. 5-16. (2013), 査読有
9. Chujo, K., Anthony, L., Oghigian, K., and Uchibori A.: "Paper-Based, Computer-Based, and Combined Data-Driven Learning Using a Web-Based Concordancer" Language Education in Asia 3(2). 132-145 (2012), 査読有
10. Nakajo, K., Nishigaki, C., Uchiyama, M. Anthony, L.: "二言語コンコーダンサーWebParaNews と AntPConc を利用した DDL 授業の実践" Proceedings of the Japan Association for English Corpus Studies (JAECS) Annual Conference. (2012), 査読有
11. Nakajo, K., Nishigaki, C. Anthony, L.: "日英パラレルコーパス検索サイトの公開: 開発と実践利用" Proceedings of the 53nd Annual National Conference of the Japan Association for Language Education and Technology. (2012), 査読有
12. Anthony, L.: "Automatic Creation of Academic Vocabulary Lists and Example Sentences for Science and Engineering students" Proceedings of the 41st Annual Conference of the English Language Education Society of Japan. 7-14 (2012), 査読有
13. Anthony, L.: "The Waseda University CELESE Program: A Large-Scale, Centralized ESP Program for Scientists and Engineers" Proceedings of 理工系英語教育を考える. 45-49 (2012), 査読有
14. Anthony, L.: "Products, processes, and practitioners: A critical look at the importance of specificity in ESP" Taiwan International ESP Journal 3(2). 1-18 (2012), 査読有
15. Anthony, L.: "Identification and Automatic Correction of Common Article Errors in Asian Learner Writing" Proceedings of the Asia Pacific Corpus Linguistics Conference. (2012), 査読有
16. Anthony, L.: "The Waseda University CELESE Program: A Large-Scale, Centralized ESP Program for Scientists and Engineers" 理工系英語教育を考える論文集. 39-45 (2012), 査読有
17. Anthony, L.: "Automatic Creation of Academic Vocabulary Lists and Example Sentences for Science and Engineering students" 日本英語教育学会 第 41 回年次研究集会 論文集. 7-14 (2012), 査読有
18. Anthony, L.: "Identification and Automatic Correction of Common Article Errors in Asian Learner Writing" Proceedings of the Asia Pacific Corpus Linguistics Conference (APCL 2012). 25-27 (2012), 査読有
19. Anthony, L.: "Products, processes, and practitioners: A critical look at the importance of specificity in ESP" Taiwan International ESP Journal (TIESPJ) 3-2. 1-18 (2012), 査読有
20. Anthony, L: "Introducing Corpus-Based Methods into a Large-Scale Technical Writing Program for Scientists and Engineers" Proceedings of the Corpus Linguistics Conference (CL 2011). (2011), 査読有
21. Bhatia, V., Anthony, L., and Noguchi, J: "ESP in the 21st Century: ESP Theory and Application Today" Proceedings of the JACET 50th Commemorative International Convention (JACET 50). 143-150 (2011), 査読有
22. Anthony, L., Nishina, Y., Takahashi, K., and Handford, M.: "Current Trends in Corpus Linguistics: Voices from Britain" Proceedings of the JAECS Annual Conference 2011. 12-13 (2011), 査読有
23. Anthony, L.: "Three (not so easy) Steps to Developing a Successful Large-Scale ESP Program in Asia" Proceedings of the 3rd International Conference on English for Specific Purposes in Asia (ESPA 2011). (2011), 査読有
24. Anthony, L.: "Why ESP practitioners do NOT need to be subject specialists" Proceedings of the 2011 International Conference and Workshop on English for Specific Purposes (ICESP 2011). 39-52 (2011), 査読有

〔学会発表〕（計 48 件）

1. Anthony, L. "Applications of Corpus Linguistics in Language Materials Design and In-Class Teaching and Learning" Distinguished Lecture Series given at Temple University Japan (招待講演). (2015, May). Tokyo/Osaka, Japan.
2. Anthony, L. "Applications of Corpus Linguistics in ESP Research: A Practical Guide" Invited workshop given at the 2015 International Forum on Applied Foreign Languages, National Kaohsiung University of Hospitality and Tourism (招待講演). (2015, May). Kaohsiung, Taiwan.
3. Anthony, L. "New Directions in Corpus Design, Tool Development, and Researcher Interaction" Invited lecture given at the Centre for Corpus Research Special Seminar, The University of Birmingham (招待講演). (2015, March). Birmingham, UK.
4. Anthony, L. "Analyzing Corpora with AntConc: From Basics to Best Practices" Invited workshop given Corpus Research Day 2015, The University of Cardiff (招待講演). (2015, March). Cardiff, UK.
5. Anthony, L. "New Developments in Corpus Tools for Data Collection, Analysis, and Visualization" Invited lecture given at The University of Nottingham (招待講演). (2015, February). Nottingham, UK.
6. Anthony, L. "Corpus Tools: Past, Present, and Future" Ertegun invited lecture given at Oxford University (招待講演). (2015, February). Oxford, UK.
7. Anthony, L. "A Hands-On Introduction to AntConc: Working with DIY corpora" Invited lecture for the Oxford University, IT Services, Corpus Linguistics course given at Oxford University (招待講演). (2015, February). Oxford,

UK.

8. Anthony, L. and Baker, P. "Automated prototypical text detection for corpus and critical discourse studies using KeyAnt" UCREL Corpus Research Seminar given at Lancaster University (招待講演). (2015, January). Lancaster, UK.
9. Anthony, L. "New AntLab Corpus Tools for English Language Researchers, Teachers, and Learners" Invited lecture given at The University of Huddersfield (招待講演). (2014, October). Huddersfield, UK.
10. Anthony, L. "Working with the AntConc Corpus Tool: A Guide For Teachers (and Learners)" Invited workshop given at the Southern University of Science and Technology (STUST) (招待講演). (2014, September). Tainan, Taiwan.
11. Anthony, L. "A View to the Future in Corpus Tools Development" Plenary speech given at the 11th Teaching and Language Corpora Conference (TALC 11), Lancaster University (招待講演). (2014, July). Lancaster, UK.
12. Anthony, L. "Introducing Corpora and Corpus Tools into the Technical Writing Classroom" Invited workshop given twice at the Summer Institute for Creative and Discovery-based Approaches to University Undergraduate Discipline-Specific Writing Programmes (招待講演). (2014, May). City University of Hong Kong, Hong Kong.
13. Anthony, L. "New desktop and web-based parallel concordance tools for corpus linguists" UCREL corpus research seminar given at Lancaster University (招待講演). (2014, May). Lancaster, UK.
14. Anthony, L. "AntPConc: A Freeware Multi-Platform Parallel Concordancer" Paper presented at the American Association for Corpus Linguistics (AACL 2014). (2014, September). Flagstaff, Arizona, US.
15. Chujo, K., Mizumoto, A., Oghigian, K., Anthony, L., and Nishigaki, C. "Comparing DDL and Non-DDL for Different Student Learning Styles" Poster presented at the American Association for Corpus Linguistics (AACL 2014). (2014, September).Flagstaff, Arizona, US.
16. Anthony, L. "Corpus Tools Brainstorming Session" Workshop given at the American Association for Corpus Linguistics (AACL 2014). (2014, September). Flagstaff, Arizona, US.
17. Anthony, L., Chujo, K., Yokota, K. and Mizumoto A.: "Broadening the Scope of Parallel Corpus Tools: Using AntPConc in the DDL Class" Second Asia Pacific Corpus Linguistics Conference (APCLC 2014). (2014, March). The Hong Kong Polytechnic University, Hong Kong
18. Anthony, L. and Nation, I.S.P.: "Freeware Vocabulary Profile and Simplification Tool for Mid-Frequency Reader Creation" Vocab@Vic Conference. (2013, December). Victoria University of Wellington
19. Anthony, L. Burd, A.: "A novel approach to medical program assessment using vocabulary profiling" Vocab@Vic Conference. (2013, December). Victoria University of Wellington
20. Chujo, K., Anthony, L., and Nishigaki, C.: "パラレルコーパスを活用する英語授業の実践: フリーウェア WebParaNews と AntPConc を使ってみる" JACET Kanto 7th Annual Conference. (2013, June). Aoyama Gakuin University
21. Anthony, L.: "From model building to corpus analysis to ESP materials creation: A three-step procedure with application in mathematics research article writing instruction" International Symposium on Innovative Teaching and Research in ESP 2014(招待講演). (2014, February). University of Electro-Communications
22. Anthony, L.: "AntConc in Action: Using Corpus Linguistics Tools and Techniques to Investigate Morphology, Syntax, Semantics, Pragmatics, and Language Variation" 2nd Korea Association of Corpus Linguistics Conference (招待講演). (2013, December). Korea University
23. Anthony, L.: "Developing Effective International Communication Skills: From Localized to Globalized Norms" 2nd International Conference of the Chinese Association for ESP and The 5th International Conference on ESP in Asia (招待講演). (2013, December). Fudan University
24. Anthony, L.: "Corpus-Based Explorations of Discourse in Language and Literature" Hwa Kang International Conference on English Language and Literature (招待講演). (2013, May). Chinese Culture University
25. Anthony, L.: "Developing AntConc for a new generation of corpus linguists" Corpus Linguistics Conference (CL 2013). (2013, July). Lancaster University
26. Anthony, L.: "Easifying KWIC Concordance Lines: The Case for Vocabulary/Range-Level Sorting" The American Association for Corpus Linguistics. (2013, January). San Diego State University, San Diego, US.
27. K. Chujo, Anthony, L. and K. Oghigian: "Using AntPConc to Teach Remedial Grammar. The American Association for Corpus Linguistics" The American Association for Corpus Linguistics. (2013, January). San Diego State University, San Diego, US.
28. Anthony, L. and Bowen, M.: "The language of mathematics: A corpus-based analysis of research writing in a neglected field" Joint International Conference of The 1st International Conference of the Chinese Association for ESP and The 4th International Conference on ESP in Asia. (2012, December). The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong.
29. Anthony, L.: "Empowering students in the English language classroom through corpus tools and data-driven learning (DDL)" A special invited lecture at Tsuda College, Tokyo, Japan (招待講演). (2012, December). Tsuda College, Tokyo, Japan
30. Anthony, L.: "Understanding Writing and Oral Presentation English in Science and Engineering: A Scientific Analysis" A special invited lecture at Hsinchu, Taiwan: National Chiao Tung University (招待講演). (2012, November). National Chiao Tung University, Hsinchu, Taiwan
31. Anthony, L.: "Designing software for multi-platform, multi-lingual audiences: The case of AntConc" IEEE Professional Communication

Society - Japan Chapter Annual Conference. (2012, October). (2012, October). The University of Aizu, Aizu Wakamatsu, Japan.

32. Anthony, L.: "The Past, Present, and Future of Software Tools in Corpus Linguistics" The International Conference of Korea Association of Corpus Linguistics (招待講演). (2012, October). Waseda University, Tokyo, Japan.
33. Anthony, L.: "Practical Guide to Using Corpus Linguistics in Research and the Classroom" A two-day workshop on introductory corpus linguistics at Fudan University, Shanghai, China (招待講演). (2012, September). Fudan University, Shanghai, China
34. Anthony, L.: "Advances in Corpus Informed ESP Research and Teaching. A Practical Guide to Teaching ESP Using Data-Driven Learning (DDL) Tools and Techniques" ESP Symposium(招待講演). (2012, September). NAIST, Nara, Japan
35. Anthony, L.: "Applications of corpus linguistics in language teaching and research" JALT Kyoto Chapter (招待講演). (2012, July). Campus Plaza Kyoto, Kyoto, Japan
36. Anthony, L.: "Understanding Character Encodings: The first (and most important) step to handling non-English corpora" Statistics, Corpora and Language Learning Workshop (招待講演). (March 8, 2012). Tokyo, Japan: Waseda University
37. Anthony, L.: "Teaching with AntConc: コーパスツールを使用したテクニカルライティング指導の実践ガイド [Teaching with AntConc: Practical guide to using corpus tools in the technical writing classroom" 42nd Conference of The English Language Education Society of Japan (JELES 42). (2012, March). Tokyo, Japan: Waseda University
38. Anthony, L.: "Identification and Automatic Correction of Common Article Errors in Asian Learner Writing" Asia Pacific Corpus Linguistics Conference (APCL 2012). (2012, February). Auckland, NZ: University of Auckland
39. Anthony, L.: "Three (not so easy) Steps to Developing a Successful Large-Scale ESP Program in Asia" 3rd International Conference on English for Specific Purposes in Asia (ESPA 2011) (招待講演). 2011, November). Xi'an, Shaanxi, P. R. China: Xi'an Jiatong University
40. Anthony, L.: "Applications of Corpus Linguistics in ESP Research and Teaching" 3rd International Conference on English for Specific Purposes in Asia (ESPA 2011) (招待講演). (2011, November). Xi'an, Shaanxi, P. R. China: Xi'an Jiatong University
41. Anthony, L., Nishina, Y., Takahashi, K., and Handford, M.: "Current Trends in Corpus Linguistics: Voices from Britain" JAECS Annual Conference 2011 (招待講演). (2011, October). Kyoto, Japan: Kyoto University of Foreign Studies
42. Anthony, L.: "Why ESP practitioners do NOT need to be subject specialists" 2011 International Conference and Workshop on English for Specific Purposes (ICESP 2011) (招待講演). (2011, October). Taichung, Taiwan: HungKuang University
43. Anthony, L.: "An Introduction to Corpus Linguistics for ESP Practitioners" 2011 International Conference and Workshop on English for Specific Purposes (ICESP 2011) (招待講演). (2011, October). Taichung, Taiwan: HungKuang University
44. Bhatia, V., Anthony, L., and Noguchi, J: "ESP in the 21st Century: ESP Theory and Application Today" JACET 50th Commemorative International Convention (JACET 50) (招待講演). (2011, August). Fukuoka, Japan: Seinan Gakuin University
45. Anthony, L., Naerssen, M., Westerfield, K.: "2012 Workshop on English for Specific Purposes: Theory and Application" Taiwan ESP Society Seminar (招待講演). (2011, July). Taiwan
46. Anthony, L.: "Introducing Corpus-Based Methods into a Large-Scale Technical Writing Program for Scientists and Engineers" Corpus Linguistics Conference (CL 2011). (2011, July). Birmingham, UK.
47. Anthony, L.: "A, An, and The: Automatically Identifying and Correcting the Most Common Errors in English Article Usage" JaltCALL 2011 Annual Conference. (2011 June). Kurume, Japan: Kurume University
48. Anthony, L.: "Introduction to Corpus Linguistics for Japanese Language Instructors" Institute for Digital Enhancement of Cognitive Development (DECODE) Workshop (招待講演). (2011, April). Tokyo, Japan: Waseda University

６．研究組織

(1)研究代表者

アントニ　ローレンス（Laurence Anthony）

早稲田大学　理工学術院　教授

研究者番号：10258204

〔その他〕

ホームページ等

http://www.laurenceanthony.net/

http://www.laurenceanthony.net/software.html