

EFL 授業におけるライティング分析支援ツールの開発

橋本 隆子^{1,a)} 山内 真理¹

概要：従来の日本の EFL(English as a Foreign Language) 授業では、「読む、聞く」といったインプット活動に重点が置かれ、「書く、話す」といったアウトプット活動が不足する傾向にある。著者(山内)はこれまでアウトプット活動の一環としてブログを利用したライティング指導を行ってきたが、指導の効果を測定するためには、学生のライティング結果を客観的に分析する必要がある。そこで我々は学生によるブログの書き込みを入力とし、学習効果測定を支援するための各種情報(語数、語彙数、文数、単語頻度等)を出力する「ライティング分析支援ツール」の開発を行った。本稿では、その概要、出力情報、分析例を紹介し、アウトプット活動の学習効果測定における本ツールの有効性について考察する。

キーワード：EFL 授業、アウトプット活動、ライティング指導、ブログ、分析支援ツール

1. はじめに

外国語を学ぶ際には、「読む、聞く」といったインプット活動と同様に「書く、話す」といったアウトプット活動も重要となる [1]。しかしながら、従来の日本の EFL (English as a Foreign Language) 授業では、カリキュラムの制約や授業時間の制限等により、インプット活動に重きがおかれ、アウトプット活動が不足する傾向にあった [2]。

著者(山内)はこれまでアウトプット活動の一貫として、ブログを利用したライティング指導を行ってきた。近年、インターネット上での情報発信が手軽に行えるようになり、ブログ等のソーシャルメディアでの発信・共有活動を取り入れた外国語学習も一般的になってきている [3]。アウトプット活動に対する抵抗感(恥ずかしい、不安であるなど)も低減でき、「英作文レポート」等の課題に比べて、自然に作成できることも特徴である。

一方で、ブログとして提出されたアウトプットをベースに学生のライティングスキルを指導していくためには、きめ細かい評価・フィードバックが必要となり、指導者側の負担が大きくなるという課題がある。指導者の負担を軽減するためには、指導に役立つ情報が容易に得られ、学習効果を評価分析できる環境が必要となる。

そこで我々は、学生によるブログ投稿を入力とし、学習効果測定のための各種情報(語数、語彙数、文数、単語頻

度等)を出力する「ライティング分析支援ツール」の開発を行った。本稿では、その概要、出力情報、分析例等について紹介する。ライティング指導の学習効果測定における本ツールの有効性についても考察を行う。

本稿は以下のような構成になっている。第 2 章でブログを利用したライティング指導の概要について説明する。第 3 章で今回開発した「ライティング分析支援ツール」を紹介する。第 5 章で「ライティング分析支援ツール」の分析例を示し、ライティング指導の学習効果測定における本ツールの有効性について考察する。第 4 章で結論と今後の方向性について述べる。

2. ブログを利用したライティング指導

外国語学習においてアウトプットは、言いたいことと伝えることの間にあるギャップへの気づきをもち、自分の中間言語の仮説を検証する機会となる。統語的な言語処理を促すなど、習得にとって重要な役割を果たしている [1]。

しかし、日本の EFL 環境では授業外で英語のアウトプットを産出する機会が少ないのが現状である。実際、種々の教員意識調査等 [2][4] から、現時点では、授業内のアウトプット活動の経験が不足している学生が相当数存在することが推察される。著者(山内)が 2013 年度に大学 1, 2 年生のクラス 64 名に対して行った調査でも、中高の英語の授業では文法説明には時間をかけるが、文法事項を使ったアウトプット練習は少ない、穴埋めや並べ替え形式のアウトプット練習はしたが、言いたいことを表現するアウトプット活動は少なかったという傾向が見られている。言語習得にとって重要なアウトプット経験が不足している学生

¹ 千葉商科大学
Chiba University of Commerce, Ichikawa, Chiba 272-8512, Japan

^{a)} takako@cuc.ac.jp

にとって、ブログを利用したアウトプット活動は、1) 英文の課題よりも自然な伝達行為に近い形で「自分の言いたいこと」を伝えられる、2) 不慣れなアウトプット活動に対する抵抗感は想定されるものの [5], 口頭会話に比べると敷居が低い、3) 身近な話題に関して書き込みをすることにより親近感が増し、ペアワーク等の授業内活動の活発化につながるといった利点がある [6]. 2011 年度より著者 (山内) が実施しているアウトプット活動は、Blogger[7] を利用した授業用ブログを用意し、毎週決まったテーマで学生が投稿するというものである. 年度や学期により実施の細部は異なるが^{*1}, 大半の学生の英語力向上に役立つという実感を得ることができた.

ブログによるアウトプット活動の課題は、修正フィードバックと学習効果測定である. 修正フィードバックについては、2011 年度にブログ投稿へのコメントとして、明示的な修正フィードバックを与えることを試みたが、投稿内容とコメントとのバランスや所用時間など実施上の課題が残った. 学習効果測定については、2012 年度に予備調査として、受講生の英文産出スピードや陥りやすい文法エラーについて、年度初めと年度末のライティングで比較分析を行った. 課題の条件を厳密には揃えていないため参考値ではあるが、全体として、より短時間でより多くの英文を産出できるようになっており、同時に文法エラーの減少も見られることが確認できた [8][9]. 学生のライティングデータを継続的に蓄積し、こうした学習効果測定を効率よく行うことができれば、より効果的な修正フィードバックや、学習素材の提供が可能となり、学生のライティングスキルの向上に有効であると考ええる.

3. ライティング分析支援ツール

ブログによるライティング指導の学習効果測定を支援するために、学生がブログ投稿を入力とし、各種情報 (語数、語彙数、文数、単語頻度等) を出力する「ライティング分析支援ツール」の開発を行った. 本章ではその概要について述べる.

本ツールは統合開発環境 IntelliJ IDEA (ver. 11) 上で、Java (ver. 5.2) 及び Scala (ver. 2.9) を用いて開発している. ライティングデータ読み込み (ReadBLog) とライティングデータ解析 (ParseBLog) の 2 つのサブツールから構成される. 図 1 は本ツールの流れである. 以下、各サブツールについて説明する.

3.1 ReadBLog

本ツールの入力には Blogger のエクスポートファイル (XML データ) と検索用タグ *KeyTag* である. XML データ内の

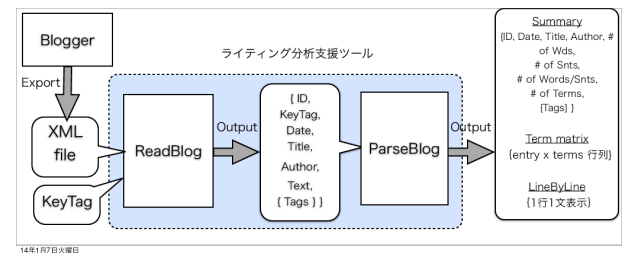


図 1 ライティング分析支援ツール概要

エントリ (投稿) 群から *KeyTag* がタグ付けされているエントリ集合 $\{entry\}$ を抽出する. 各エントリ $entry_i$ を以下のタプルからなるデータとし、CSV 形式ファイルとして出力する.

$$entry_i = \{ID, KeyTag, Date, Title, Author, Text, \{Tags\}\}$$

ここで *ID* は各エントリ $entry_i$ のエントリ ID, *Date* は投稿日付, *KeyTag* は検索用タグ, *Title* はエントリタイトル, *Author* は投稿者, *Text* は本文, $\{Tags\}$ は該当エントリに付加されたタグ集合 (*KeyTag* を含む) である. 表 1 に ReadBLog の出力例を示す.

3.2 ParseBLog

本サブツールは ReadBLog の出力結果である CSV ファイルを入力として、形態素解析を行い、Summary, TermMatrix 及び LineByLine の 3 種類のファイルを出力する. 形態素解析は、オープンソースのテキスト検索エンジンライブラリー Lucene [10] を用いて行っている. 以下、Summary, TermMatrix, LineByLine の各ファイルの仕様について説明する.

3.2.1 Summary

本ファイルは、 $entry_i$ の *Text* を形態素解析し、以下に示される値を算出し、 $\{entry\}$ 全体で 1 つの CSV ファイルとしたものである.

$$entry_i = \{ID, Date, Author, Title, Text, \#ofWds, \#ofWds/Snts, \#ofSnts, Ave.\#ofWds/Snt, \#ofTerms, \{Tags\}\}$$

ここで *ID* は $entry_i$ のエントリ ID, *Date* は投稿日付, *Author* は投稿者, *Title* はエントリタイトル, *Text* は本文, *#ofWds* は出現ワード数, *#ofWds/Snts* は文ごとのワード数 (”-”でワード数が連結して示される), *#ofSnts* は文章数, *Ave.#ofWds/Snt* は 1 文あたりの平均ワード数, *#ofTerms* は出現単語数 (出現ワードから重複を省いたもの), $\{Tags\}$ は本エントリに付加されたタグ集合

^{*1} たとえば 2013 年度後期は、Moodle での Wiki や Workshop を利用したライティング活動も取り入れ、ブログ投稿の頻度は減った

表 1 ReadBlog 出力例

ID	KeyTag	Date	Title	Author	Text	Tags
1	2B	2013/5/16	About me	** **	About me Hellow everyone my name is Arashi Ishibuchi.	AboutMe, Arashi
2	2B	2013/5/15	GW holidays	** **	Hellow everyone I'm Shota. Today I introduces my GW holidays.	Shota, GW holidays
3	2B	2013/5/15	About me	** **	Hellow everyone my name is Shota Takahashi. I'm from Iwate but I live in Ichikawa.	AboutMe, Shota
4	2B	2013/5/15	GW holidays	** **	Hello everyone!! I'm Daiki. This year GW holiday I was busy.	GW holidays, Daiki
5	2B	2013/5/15	GW holidays	** **		GW holidays, Norimitsu
6	2B	2013/5/15	About me	** **		AboutMe, Norimitsu

表 2 ParseBlog 出力例 : Summary

ID	Date	Author	Title	Text	# of Wds	# of Wds /Snts	# of Snts	Ave. # of Wds /Snt	# of Terms	Tags
1	2013/5/16	** **	About me		84	9-5-5-5-9-6-9-6-13-6-9	11	7	58	,
2	2013/5/15	** **	GW holidays		83	4-6-10-12-8-5-7-11-11-5-4	11	7	57	,
3	2013/5/15	** **	About me		66	7-8-5-6-16-8-5-7-4	9	7	53	,
4	2013/5/15	** **	GW holidays		65	4-7-6-4-5-7-7-9-9-7	10	6	50	,
5	2013/5/15	** **	GW holidays		69		12	5	52	,
6	2013/5/15	** **	About me		78		11	7	55	,

(KeyTag を含む) である。表 2 に出力例を示す。

3.2.2 TermMatrix

本ファイルは、エン트리集合 $\{entry\}$ と出現単語の行列である。エン트리 ID (ID)、投稿日付 ($Date$)、投稿者 ($Author$)、タイトル ($Title$)、語彙数 ($\#ofTerms$)、及び各単語の出現頻度が記載される。表 3 に出力例を示す。

3.2.3 LineByLine

本ファイルは、 $entry_i$ の $Text$ を文単位で分割し、1 行に 1 文ずつ表示したものである。上記 2 つのファイルと同様、 $\{entry\}$ 全体で 1 つの CSV ファイルとなる。

$entry_i = \{ \{ID, Date, Author, Title, Snt, SntID, \#ofWds/Snt\} \}$

ここで ID は $entry_i$ のエン트리 ID、 $Date$ は投稿日付、 $Author$ は投稿者、 $Title$ はエン트리タイトル、 Snt は 1 文、 $SntID$ は文番号 (1 から順に振られる)、 $\#ofWds/Snt$ は該当文の出現ワード数である。表 4 に出力例を示す。このファイルにより学生の投稿を 1 文ずつ確認することが容易となる。

4. ライティング分析支援ツールの分析例

今回の分析対象は以下のデータである。

- 投稿者：大学生 20 名。
- テーマ：表 5 に示されている 6 つのテーマ。各テーマがブログエントリのタイトルとなる。
- 投稿期間：2013 年 4 月 24 日～12 月 23 日。テーマごとに投稿期間が異なる。表 5 参照。
- 総投稿数：101 件。テーマごとに投稿数が異なる*2。表 5 参照。

上記入力データに対し、本ツールを適用し、Summary, TermMatrix, LineByLine ファイルを出力した。以下、この 3 つのファイルを基にした分析を行う。

4.1 数値変化による分析

Summary ファイルのデータを基にブログのテーマごとに以下の数値を算出する。

(1) 1 投稿あたりの平均語数

*2 テーマによって投稿数が異なるのは、重複して提出した学生や未提出の学生がいたためである。

表 3 ParseBlog 出力例 : *TermMatrix*

ID	Date	Author	Title	# of Terms	about	me	hellow	everyone	my	name	is	...
1	2013/5/16	** **	About me	58	1	1	1	2	1	6	1	1
2	2013/5/15	** **	GW holidays	50	0	0	1	1	0	1	0	0
3	2013/5/15	** **	GW holidays	52	1	0	1	2	0	0	0	0
4	2013/5/15	** **	About me	55	1	0	0	4	2	5	0	0
5	2013/5/14	** **	GW holidays	49	0	0	1	3	1	1	0	0
6	2013/5/12	** **	About Me	51	0	0	1	1	0	2	0	0

表 4 ParseBlog 出力例 : *LineByLine*

ID	Date	Author	Title	Snt	Snt ID	# of Wds / Snt
...	...	...	...	...	...	...
9	2013/5/15	** **	About me	Hellow everyone my name is ** **.	1	7
9	2013/5/15	** **	About me	I'm from Iwate but I live in Ichikawa.	2	8
9	2013/5/15	** **	About me	Now I live on my own.	3	5
9	2013/5/15	** **	About me	So I can cook a little.	4	6
...	...	...	...	...	...	...
10	2013/5/16	** **	About me	Hellow everyone my name is ** **.	1	7
10	2013/5/16	** **	About me	I am from Higashimukoujima in Tokyo.	2	5
10	2013/5/16	** **	About me	I like comics and anime.	3	5
...	...	...	...	...	...	...

- (2) 1 投稿あたりの平均語彙数
- (3) 1 投稿あたりの平均文数
- (4) 1 文あたりの平均語数

指導が進むに従い、1 投稿あたりの語数、語彙数が増えていくと予測される。また 1 投稿あたりの文数も増え、1 文あたりの平均語数も増えていくと考えられる。

表 6 は、分析対象に対して、ブログのテーマごとに上記数値を算出した結果である。図 2 から図 5 は、表 6 のデータをグラフ化したものである。今回の分析では、1 投稿あたりの平均語数 (図 2)、1 投稿あたりの語彙数 (図 3) に大きな変化は見られなかったが、1 文あたりの平均語数 (図 5) は 6.10 から 10.40 と明らかに増加している。これは 1 文をより多くの単語を用いて作成できるようになったことを示している。文の長さは長いほど良いとは言えないが、平均 6 語から平均 10 語に変化したとすれば、修飾語句や接続詞等を含むより複雑な文の増加が推察され、学生のスキル向上を確認できたと考えられる。

学生のライティングをこうしたシンプルな数値で時系列に評価することで、学習効果を測定していくことが可能となるのではないかと考える。

4.2 語彙変化による分析

次にライティング指導によって、学生の語彙がどのように変化したかを TermMatrix ファイルの「エントリ × 単語行列」を基に作成したグラフ構造によって考察する。エントリ × 単語の行列をテーマごとにグルーピングし、各テーマにおける単語の共起関係に基づいてグラフ構造を作成する。例えば、テーマ「About me」のエントリ $entry_i$ にお

いて「I am a student. My school is in Chiba.」という書き込みがあった時、TermMatrix の $entry_i$ の行では「I, am, a, student, my, school, is, in, Chiba」の各単語に対して出現頻度が ≥ 1 となる。本処理では「I, am, a, student, my, school, is, in, Chiba」のように出現頻度が ≥ 1 の単語を共起単語とみなし、それぞれをエッジで結び、無向グラフを作成する。

図 6 は、テーマ「About me」のエントリ群における共起単語のグラフであり、図 7 は、テーマ「World Heritage」のエントリ群における共起単語のグラフである。グラフの可視化は、オープンソースのグラフ可視化環境である Gephi[11] により行っている。図 6 は自己紹介の書き込みであり、「i, school, live, hobby..」といった自分を説明する単語群が見られる。一方、「World Heritage」についての書き込みである図 7 では、3 人称の Be 動詞である「is」、仮定や条件を表現する「if」、目的や方向を表現する「to」などが現れている。こうした可視化を行うことにより、語彙数が増えたということだけでなく、利用する単語のバリエーションや難易度が向上していることもわかる。語彙のバリエーションは、テーマの違いによっても影響されるが、例えば「ゴールデンウィーク」と「夏休み」のように、自分が何をしたかを描写するという似たテーマでの投稿について語彙数のバリエーションの変化を観察することも可能となる。今回開発したライティング分析支援ツールにより、英文の表現力が向上している様子を評価することができる。と考える。

4.3 エラー分析の効率化

第 2 章で学習効果測定として「文法エラー分析」に言及

したが、著者が推進しているライティング指導では、英語として正しい語順で文を組み立てられているかどうか [12] に関わるエラーを、優先的に指導・修正すべき事項として重視している。学習効果測定のためのエラー分析として、当該タイプのエラーを含む文の出現比率の比較がある。本支援ツールの出力形式の1つである LineByLine (1行1文表示。書き手、日時、テーマなどの基本的なテキスト情報も保持) を利用することで、このエラー分析が格段に容易になる。

また、平均語数の近い投稿群を比較したり、一定の長さをもつ (例えば10語以上の) 文のみを抽出した上で、含まれる統語上のエラーを時系列で比較することも可能になり、本ツールがエラー分析における指導者の負荷を大幅に軽減できると考える。

表5 入力データ

ID	テーマ名	内容	投稿期間	投稿数
I	About me	自己紹介	2013年4月下旬～5月上旬	21
II	GW holidays	GWの思い出	2013年5月中旬～下旬	19
III	Sharing food photos	食べ物の写真をSNSでシェアするかについて	2013年6月中旬～7月上旬	12
IV	What I usually do on weekends	週末の過ごし方	2013年7月上旬～中旬	17
V	Summer Vacation	夏休みの思い出	2013年9月下旬～10月上旬	17
VI	World heritage	世界遺産について	2013年10月中旬～10月下旬	15

表6 ライティング結果の各種数値変化

ID	投稿数	平均語数 / 投稿	平均語彙数 / 投稿	平均文数 / 投稿	平均語数 / 文
I	21	90.95	60.76	13.95	6.19
II	19	87.84	57.53	11.58	7.47
III	12	74.33	41.83	8.08	9.08
IV	17	54.06	39.29	7.65	6.71
V	17	105.29	65.82	11.53	9.47
VI	15	71.27	44.93	6.87	10.40

5. おわりに

EFL 授業におけるアウトプットの重要性に注目し、ブログを利用したライティング指導の効果測定を目的としたライティング分析支援ツールを開発した。

本ツールは、学生によるブログ投稿を入力とし、学習効果測定を支援する各種情報 (語数、語彙数、文数、単語頻度行列等) を出力する。本稿では、ツールの概要、出力情報、分析例を紹介し、こうしたツールがライティング指導

1投稿あたりの平均語数

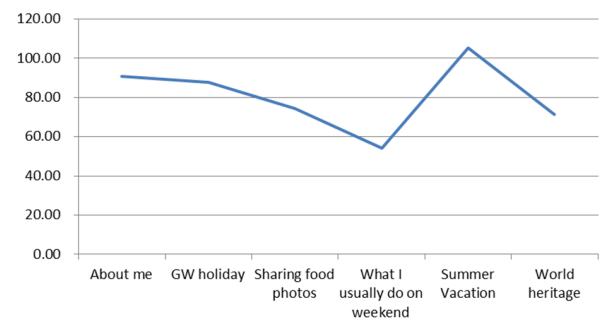


図2 1投稿あたりの平均語数

1投稿あたりの平均語彙数

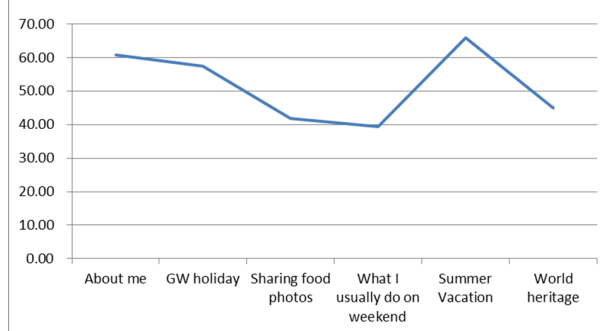


図3 1投稿あたりの平均語彙数

1投稿あたりの平均文数

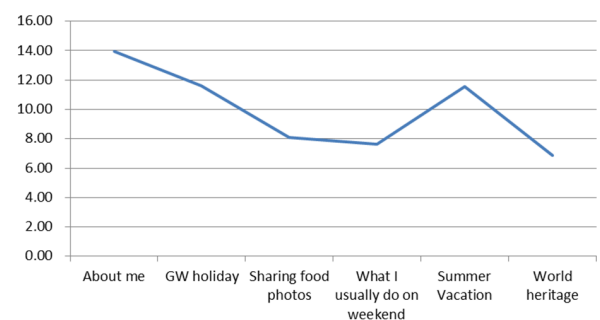


図4 1投稿あたりの平均文数

の支援に役立つことを確認した。

今後は、ライティング指導の課題の一つである「エラー分析」にさらに注力し、指導者の負荷を軽減するインタフェースの構築を考える。それとともに、ブログデータを継続して蓄積し、各種分析を行い、学習効果測定のためにこういった情報が必要となるかをさらに考察していく。

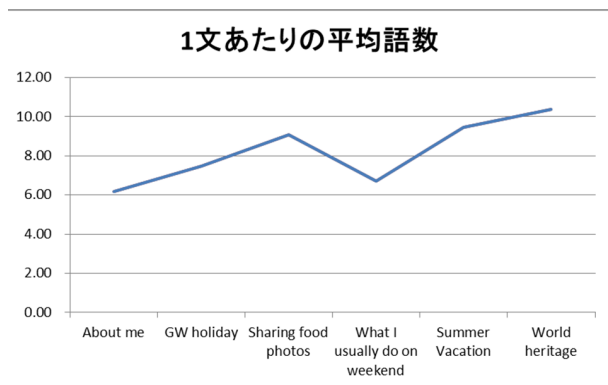


図 5 1 文あたりの平均語数

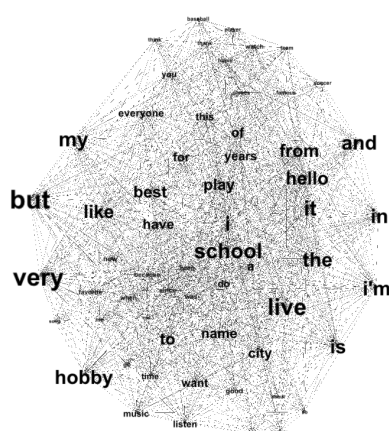


図 6 トピック「About me」の共起単語群

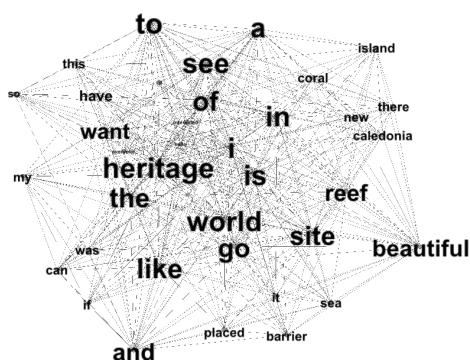


図 7 トピック「World Heritage」の共起単語群

参考文献

- [1] Swain, M., Gass, S. M., Madden, C.: *Communicative competence: Some roles of comprehensible input and comprehensible output in its development*, Input in Second Language Acquisition, Input in second language acquisition, No. Cambridge, pp.235-256 (1985).
- [2] Champbell C., Kikuchi A. and Palmer, R.: *Bridging the Gap Between CLT Theory and Practice in a Japanese Junior High School*, 関西大学外国語教育研究 12, 2006-10,

- pp.1-20 (2006).
- [3] Thomas, M. (Ed). : *Handbook of Research on Web 2.0 and Second Language Learning* Information Science Reference (2009).
- [4] 坂本優美子: 日本の公立高校英語教師の CLT と伝統的な教授法に対する態度, 摂南大学教育研究, 1, pp.53-74 (2005) .
- [5] Doyon, P.: *Shyness in the Japanese EFL Class: Why It Is a Problem, What It Is, What Causes It, and What to Do About It*, The Language Teacher, 24(1),pp.12-16, 37 (2000).
- [6] Yamauchi, M.: *Effective Implementation of a Class Blog in the Traditional Classroom Setting: Toward More Learner-centered EFL Instruction*, GloCALL 2011, Manila (2011).
- [7] Google: Blogger (ブログ作成・公開ツール), 入手先 <<http://www.blogger.com/>> (2013.12.28).
- [8] 山内 真理, 橋本 隆子: 日本の EFL 授業における意味重視のアウトプット活動, 日本リメディアル教育学会関西支部大会 2013, 大阪 (2003).
- [9] Yamauchi, M., Hashimoto, T.: *Effective Implementation of Meaning-focused Output Activities for Japanese EFL Learners*. JALTCALL 2013, Matsumoto (2013).
- [10] Apache: Lucene, 入手先 <<http://lucene.apache.org/core/>> (2013.12.28).
- [11] Gephi - makes graphs handy, 入手先 <<http://oss.infoscience.co.jp/gephi/gephi.org/>> (2013.12.28).
- [12] 山内真理, 内田充美: 日本人英語学習者の中間言語にみられる L1 の痕跡. 千葉商大紀要 49(1), 43-55 (2011).