

チャンキング過程を考慮したスラッシュ・リーディング用文書の生成

田中, 省作
九州大学情報基盤センター

山本, 祥平
九州大学大学院システム情報科学府

富浦, 洋一
九州大学大学院システム情報科学研究院

<https://doi.org/10.15017/4783573>

出版情報 : 九州大学情報基盤センター年報. 4, pp.1-8, 2004-03. 九州大学情報基盤センター
バージョン :
権利関係 :

チャンキング過程を考慮したスラッシュ・リーディング用文書の生成 Generation of Texts for Slash Reading Based on Chunking Process

田中 省作†
Shosaku Tanaka†

山本 祥平†
Shohei Yamamoto†

富浦 洋一*
Yoichi Tomiura*

† … 九州大学情報基盤センター

† … Computing and Communications Center, Kyushu University

‡ … 九州大学大学院システム情報科学府

‡ … Graduate School of Information Science and Electrical Engineering, Kyushu University

* … 九州大学大学院システム情報科学府

* … Faculty of Information Science and Electrical Engineering, Kyushu University

要旨 スラッシュ・リーディングとは学習者が適当な意味のかたまりごとに“/”（スラッシュ）が挿入された英文を素早く読み進める学習法で、学習者は英語本来の語順で英文を理解することを習慣づけることができる。しかし、安価に利用できるスラッシュ・リーディング用教材は少なく、英語教師が各自で作成しているのが現状である。そこで本稿では、英文に対してその依存構造から想定されるチャンキング過程（文中の部分表現の意味を構成する過程）に基づき、スラッシュの入れ方を決定する手法を提案する。

Abstract Slash reading is one of popular ways for learning English, that helps non-native learners understand an English sentence in a sequence order as its words appear. Though it is needed for the learners to read a lot of English sentences which have some slashes in the boundary between sense groups, there are a few texts at the moment. This paper proposes a method for inserting slashes into an English sentence automatically based on chunking process derived from its dependency structure and shows the experimental result using the proposed method.

1 はじめに

近年注目されている英語学習法の1つにスラッシュ・リーディング (slash reading)¹がある。スラッシュ・リーディングとは学習者が、英文中に適当な意味のかたまりで“/”（スラッシュ）が挿入された英文を素早く大量に読み進める学習法である。例えば、次のようなスラッシュ付きの英文

he said / many people in other Asian countries /
suffered grtealy / during World_War_II

では、スラッシュで挟まれる区間（本稿ではこれをセグメントとよぶ），“he said”に対して「彼は（なにかを）言った」、 “many people in other Asian countries”

に対して「他のアジア諸国の多くの人々が」、以後同様に「とても苦しんだ」、「第二次世界大戦中に」とセグメントごとに意味を大雑把に捉え、並行してそれらの意味を適格に結びながら文意を読み取るのである。これを繰り返すことで、英語本来の語順で英文を直接理解することが習慣づけられ²、通訳養成などでも広く用いられている。しかし、現状では自由に利用できる安価なスラッシュ・リーディング用の英文教材は少なく、英語教師が各自で準備する他ない。また、この学習法は学習者がある程度の量の英文を読解することが大切であり、多読的な要請もある。そのような意味で学習者自身が興味ある英文書を自由に選び、それからスラッシュ付きの英文書を作成し、自主的に読み進

¹その他にも、フレーズ・リーディングやチャンク・リーディングと呼ばれることもある。

²このような訓練はリーディングだけでなくリスニング力の向上にも非常に効果的であることが知られている

めるような環境があればなお望ましい。

そこで本稿では、英文に対してその依存構造から想定されるチャンキング過程を考慮してスラッシュを挿入する手法を提案する。文読解時のチャンクとは読者の短期記憶内に記憶される文中の部分表現の意味であり、「結合」「忘却」「分解」などの操作により文意に相当するチャンクを構成する過程はチャンキング過程とよばれる。本稿では「分解」が不要なチャンキング過程において、スラッシュを「ここで適当な意味にまとめる」という情報を読者に明示する記号と捉え、チャンクの結合操作が起きる語の直後スラッシュを入れる候補とし、チャンク間の区間(セグメント)長に対する適当な上限を設定する。さらに、依存関係がスラッシュがセグメント内では密に、セグメント間は疎となるように挿入する。

本手法はスラッシュ・リーディング用教材の自動生成に用いることを念頭に置いており、(i) 学習データを必要としないこと、(ii) 文に関する情報は依存構造のみであること、という特徴がある。英文に対するスラッシュの入れ方の基準は、母語話者でも一定ではなく、さらに習熟度によっても大きく変わる可能性があり、一般に大量の学習データ(スラッシュ付き英文)は期待できない。また、現在の自然言語処理技術の水準では、部分表現の統語範疇と下位の統語構造を完全に求める統語解析は十分な精度は得られていない。しかしながら、語間の依存関係であれば比較的現実的な精度が得られる可能性が高い。

2 チャンキング過程

2.1 文読解とチャンク

文読解とは読み進めた位置までの文の部分表現を短期記憶内で逐一適当な抽象的表象(意味)へ変換する作業と見なすことができる。さきの例文で、“he”まで読んだ時点では、読者の短期記憶(short-term memory)には“he”そのものではなくある特定の「彼」を連想し、次の“said”まで読んだ時点では“said”という字面ではなく「言った」という意味を構成、さらに短期記憶内の「彼」と結合して「彼は(なにかを)言った」という意味を構成する。このように読解時に構成される意味をチャンク(chunk)とよぶ³。また、「彼」と

³一般的にチャンクを議論すると、人間の認知活動において、認知対象の最も基本的な単位のことをいう。例えば、“092096”という数列をそれぞれ“0”, “9”, “2”, “...”という具合に認知すればそれは6チャンクということになる。しかしながら、同じ数列でも“092”(福岡の局番)、“096”(熊本の局番)という具合に適当に

「言った」を結合して「彼は(なにかを)言った」を構成するチャンク間の「結合」、その他チャンクの「忘却」、チャンクの「分解」といったチャンクに対する操作をチャンキング(chunking)とよぶ。つまり、文読解とはチャンキング過程であり、最終的には文意に相当するチャンクを求めることである。

2.2 教材作成におけるチャンキング過程の一つの考え方

文読解時のチャンキング過程とは、ある一定の条件のもとで文の部分表現から得られるチャンクを「結合」「忘却」「分解」などの操作によって、よりコンパクトなチャンクを構成することである。文読解時「結合」すべきチャンクは表層レベルにおいて語間の依存関係として投影されている。さまざまなチャンキング過程を考えることができるが、ここでは次のような「分解」が不要なチャンキング過程を採用する。ここで考えるべきチャンキング過程というのは、英文にスラッシュをいれるべき箇所を決定するための最も基本的な情報、すなわちお手本となるものである。よって、最も効率よく読解した場合を想定すべきである、と考えた。

2.2.1 記法と諸定義

文 s の i 番目の語を w_i とし、語間の依存構造を仮定する。 D は語のインデックスの直積の部分集合で語の依存関係を表し、 $(i, j) \in D$ ならば語 w_i に w_j が係っている。特に w_i に係る語のインデックス集合を $D(i)$ で略記する。 $D^+(i)$ を次のように定義する。

1. $j \in D(i)$ ならば $j \in D^+(i)$ である。
2. $j \in D^+(i)$ かつ $k \in D(j)$ ならば $k \in D^+(i)$ である。
3. 1, 2 を有限回適用して得られるものが $D^+(i)$ である。

主要語が w_i のチャンクを C_i とすると、 w_p まで読み進んだときにチャンク C_i が飽和している(saturated)とは、

$$D^+(i) = \emptyset \text{ または } \max D^+(i) < p$$

が成立している場合である。つまり、語 w_i が飽和しているとは、 w_i に係る語が全て出現しており、なおかつ係る語も飽和している場合である。

意味ある単位で結合して認知すれば2チャンクとなる。

i	$D(i)$	$D^+(i)$
2	1, 9	1, 4, 9, 10, 11, 12
4	3, 5	3, 5, 6, 7, 8
5	8	6, 7, 8
8	6, 7	6, 7
9	4, 10, 11	4, 10, 11, 12
11	12	12

表 1: 依存関係 D と D^+

例 1 次の文

he₁ said₂ many₃ people₄ in₅ other₆ Asian₇ countries₈
suffered₉ greatly₁₀ during₁₁ World_War_II₁₂

に対する依存構造は図 1 のようになる。このとき、 $D(i), D^+(i)$ は表 1 のようになる（ただし、表に明記しなかった $D(i), D^+(i)$ は空である）。

例えば、 $w_2(= \text{said})$ を主要語とするチャンク C_2 は $w_{12}(= \text{World_War_II})$ まで読み進めないと飽和しない。 $w_4(= \text{people})$ を主要語とするチャンク C_4 は $w_8(= \text{countries})$ まで読み進めれば飽和していると言える。□

2.2.2 「分解」が不要なチャンキング過程

語 w_p まで読み進んだときの短期記憶に蓄えられたチャンクの集合を $M(p)$ で表す。 $M(0) = \emptyset$ で、 w_p からさらに w_{p+1} を読んだ後の短期記憶の状態 $M(p+1)$ は次のように計算される。

1. T を $M(p)$ とする。
2. w_{p+1} から C_{p+1} を構成し T に加える。
3. $C_i \in T$ の主要語 w_i に対して飽和しているチャンク $C_j \in T$ の主要語 w_j が係っている、つまり、 $j \in D(i)$ ならば、 C_j を C_i に結合し、 C_j を T より削除する。
4. 3 を繰り返し、収束した T が $M(p+1)$ である。

w_p まで読んだ時点で短期記憶に蓄えられるチャンクの数はちょうど、係り先の語が出現していない、または飽和していないチャンクの数となる。以上のチャンキング過程では、チャンクが飽和するまで結合操作を行わないので、一貫して分解操作は起きない。その分、留保しておくチャンクの数が増大するので、短期記憶への負荷が大きくなる。

人の短期記憶容量は 7 ± 2 程度であるという仮説（マジカルナンバー 7 ± 2 ）[5] が知られている。文献[11]によると、実際に EDR 英語コーパス [1] の英文約 12 万

文に対して調査を行い、一文中の各語の位置で最大のチャンク数をその文のチャンク数と考えると、7 を超えるものが 1.5%、9 を超えるものが 0.01% であった⁴。多くの場合マジカルナンバー内に収まるからと言って、チャンキング過程が適当である、ということを保証するものではないが、その必要条件の一つであり、少なくともこれは満たすものである。

例 2 さきの例文のチャンキング過程は表 2 のようになり、最も短期記憶に負荷がかかるのが “Asian” を読んだ直後で、チャンク数は $5(= |M(7)|)$ となる。□

3 スラッシュの入れ方

3.1 チャンク数推移に対する制約

従来、セグメントに相当する「意味のかたまり」を定義し、そこからスラッシュ挿入の規則を間接的に与えていた [12]。本稿では、このようなスラッシュをセグメントの境界であるという考えだけでなく、読解時に「ここで適当に意味をまとめる」という情報を明示する記号とも考えることにする。すると、スラッシュを入れる箇所候補は語 w_{p-1} から w_p を読み進めたにもかかわらずチャンク数が変わらない、または減少する

$$|M(p-1)| \geq |M(p)|$$

という語 w_p の直後ということになる。するとスラッシュが入った箇所は次のような制約を充たさなければならない。

制約 1 文長 n の文 $s = w_1 w_2 \dots w_n$ に対して K 本のスラッシュを入れたとき、 $SP(k)$ を文頭から k 番目のスラッシュの位置（何番目の語の直後か）とする。スラッシュの入った箇所では次式

$$|M(SP(k)-1)| \geq |M(SP(k))| \quad (1 \leq k \leq K) \quad (1)$$

を充たす。ただし、 $SP(0) = 0, SP(K+1) = n$ である。□

例 3 表 2 でチャンク数の推移を見ると、スラッシュを入れる箇所候補は、次の ○ で囲った 5 箇所である。
he said ① many people ② in other Asian countries
③ suffered ④ greatly ⑤ during World_War_II □

⁴EDR 英文コーパスの依存構造には、and や or といった並列句、be 動詞など幾つかの文法現象については我々の直感とは合わない。これらの依存構造を修正すれば、以上の値はもっと改善される。

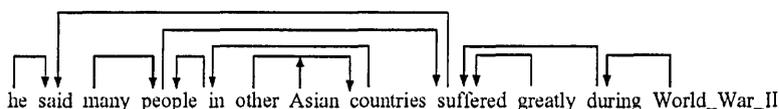


図 1: 依存構造

p	$M(p)$	$ M(p) $
1	“he”	1
2	“he said”	1
3	“he said”, “many”	2
4	“he said”, “many people”	2
5	“he said”, “many people”, “in”	3
6	“he said”, “many people”, “in”, “other”	4
7	“he said”, “many people”, “in”, “other”, “Asian”	5
8	“he said”, “many people ~ countries”	2
9	“he said”, “many people ~ suffered”	2
10	“he said”, “many people ~ greatly”	2
11	“he said”, “many people ~ greatly”, “during”	3
12		0

表 2: チャンキング過程の例

カンマについて

本手法ではさらに次のように“,”(カンマ)が付随した w_p で、なおかつ (1) 式が成立する場合は、“ w_p / w_{p+1} ”とスラッシュを入れた。カンマの用法の1つとして、主文と導入節・句の区切りといった大きな意味の境目に入ることが知られているからである [3]。

では、2番目のセグメント長 (“many ~ greatly”) は $8 \leq \theta$ となる。よって、以上のスラッシュの入れ方は不適當で、さらに ② ~ ④ の少なくとも1箇所にはスラッシュを入れる必要がある。 □

ここでは、セグメント長は単純にセグメント内の語数としてしてること、およびその上限 θ の決め方については議論の余地がある。4.3節で改めて言及する。

3.2 セグメント長に対する制約

長過ぎるセグメントはチャンキングを複雑にし、文読解失敗の原因となる。そこで、各セグメント長に対して適当な上限 θ を設ける。

制約 2 文長 n の文 $s = w_1 w_2 \dots w_n$ に対して K 本のスラッシュを入れたとき、各セグメントは次式

$$SP(k) - SP(k-1) \leq \theta \quad (1 \leq k \leq K+1) \quad (2)$$

を充たす。 □

例 4 $\theta = 7$ としよう。例 3 で、①と⑤のみのスラッシュの入れ方

he said / many people in other Asian countries suffered greatly / during World_War_II

3.3 依存関係に基づく優先順位付け

複数のスラッシュの入れ方に対する優先順位を与える。本稿では、セグメントの意味のまとめやすさとして依存関係に着目し定量化、優先順位を与えることにする。

語間の依存関係は、セグメントを意味へまとめる際の本質的な情報である。例えば、語の多義性の解消では依存関係を結ぶ語がなにかということが非常に重要となる。よって、セグメント内における依存関係は密に、セグメント間の依存関係は疎になるようにするのが望ましい。そこで、セグメント内の依存構造の数をアドバンテージ、スラッシュを挿入したために断ち切られる後方との依存関係をペナルティと考える。こ

ここで、ペナルティとしてスラッシュより後方のみを加味するのは、それより前方のセグメントはチャンクとして適当な意味にまとめられているという前提である。

k 番目のセグメント ($k-1$ 番目のスラッシュと k 番目のスラッシュで挟まれる部分表現)

$$w_{SP(k-1)+1} w_{SP(k-1)+2} \cdots w_{SP(k)}$$

内の依存関係の数は、

$$f_k = \sum_{i=SP(k-1)+1}^{SP(k)} |\{j \in D(i) : SP(k-1) < j \leq SP(k)\}|$$

である。一方、 k 番目のスラッシュ ($w_{SP(k)}$ の直後に入れられるスラッシュ) によって断ち切られる依存関係の数は、

$$g_k = \sum_{i=SP(k-1)+1}^{SP(k)} |\{j \in D(i) : j > SP(k)\}| + \sum_{i=SP(k)+1}^n |\{j : j > SP(k) \wedge i \in D(j)\}|$$

である。今回は f_k の和、 g_k の和の比を優先付けのためのスコアとした。

$$\text{Score} = \sum_{k=1}^K f_k / \sum_{k=1}^{K-1} g_k \quad (3)$$

例 5 例 4 で、②にスラッシュを入れたとすると、
he said / many people / in other Asian countries
suffered greatly / during World_War_II

$$f_1 + f_2 + f_3 + f_4 = 1 + 1 + 4 + 1 = 7$$

$$g_1 + g_2 + g_3 = 1 + 2 + 1 = 4$$

で $\text{Score} = 7/4$ となる。一方、②ではなく③にスラッシュを入れたとすると、

he said / many people in other Asian countries /
suffered greatly / during World_War_II

$$f_1 + f_2 + f_3 + f_4 = 1 + 5 + 1 + 1 = 8$$

$$g_1 + g_2 + g_3 = 1 + 1 + 1 = 3$$

で $\text{Score} = 8/3$ となり、後者のスラッシュの入れ方が優先される。□

また、(3) が同値だったものについては、スラッシュを挿入した箇所のチャンク数の減り幅の総和

$$\sum_{k=1}^K \{|M(SP(k))| - |M(SP(k) - 1)|\}$$

が大きいものを優先する。

4 実験

4.1 方法

英文に対するスラッシュの入れ方には多様な要因が考えられ、母語話者でも揺れがある。非母語話者であればなおさらである。そこで、今回はオリジナルの英文および本手法でスラッシュ付けされた英文のうち上位 1,2,5,10,15,20 位の 6 文をランダムに並び替え被験者に提示し、「読解時に自分にとって最も理解しやすいであろうスラッシュ付きの文」を 1 つを選択するという方法で、本手法による優先順位付けとの相関について調査した。被験者は大学卒業レベルの非母語話者 3 人、選択の際時間の制限は行わず辞書も自由に利用して構わないとした。英文は EDR 英語コーパス [1] の構文解析済み 124,698 文からランダムに抽出した。

例 6 次の英文

the report observed that the figure indicated fewer hospitals were refusing to treat emergency patients. に対して本手法でスラッシュ付けされた英文のうち 1,2,5,10,15,20 位の 6 文は、図 6 のように被験者に提示される。括弧内は本手法による優先順位で、被験者には提示されない。

□

また、本手法ではセグメント長に対する上限 θ を設定する必要がある。今回は被験者ごとに θ を次のような手続きで設定した。予め EDR 英語コーパス中の適当な 10 文について、それぞれ $\theta = 5, 6, 7$ の 3 通りでスラッシュ付けした優先順位最上位の 3 文をランダムに並び替え θ は伏せて被験者に提示し、最良と思われる文を選択してもらった。以上を全 10 文に対して繰り返し、最も多く選択された θ を被験者に対する θ として設定した。今回は各々 $\theta = 5, 6, 7$ となった。

4.2 結果

被験者が選択した順位とその個数を表 3 に示す (括弧内は累積の割合を示す)。上位 2 位までは (3) 式が同値となる場合が多かったことから、上位 2 番目までを選択した場合を正解と考えた。結果上位 2 位までのスラッシュの入れ方が選ばれたのは 667 文中 515 文 (77.2%) であった。ベースラインは (甘い見積もりではあるが) ランダムに選択した場合 1/3 となるので、被験者のスラッシュの入れ方と (3) 式による相関がある程度認められたと言える。

1. the report / observed that the figure indicated / fewer hospitals / were refusing to treat / emergency patients. (5)
2. the report / observed / that the figure indicated / fewer hospitals / were refusing / to treat emergency patients. (15)
3. the report observed / that the figure indicated / fewer hospitals were refusing / to treat emergency patients. (2)
4. the report / observed / that the figure indicated / fewer hospitals / were refusing to treat / emergency patients. (20)
5. the report / observed that the figure indicated / fewer hospitals were refusing / to treat emergency patients. (1)
6. the report observed / that the figure indicated / fewer hospitals / were refusing to treat / emergency patients. (10)

図 2: 被験者に提示されるスラッシュ付き英文の例 (括弧内は優先順位で被験者には表示されない)

順位	A($\theta = 7$)	B($\theta = 5$)	C($\theta = 6$)	全体
1	122	169	74	365 (0.547)
2	61	70	19	150 (0.772)
5	22	30	8	60 (0.862)
10	19	17	2	38 (0.919)
15	20	20	1	41 (0.980)
20	5	8	0	13 (1.000)
計	249	314	104	667

表 3: 実験結果

5位以降を選択した場合について考察する。特徴的だったものは、1,2位のものを選択したものが多くの部分で共通ではあるが、唯一スラッシュの箇所が多いものを選択している場合である。以下に幾つかの典型的な場合を説明する。

- “/ that S” や “/ who S” というように補文構造の直前
補文構造は語数が大きくなることも要因の1つではあるが、むしろ「文」に対して被験者が特別視している、ということがうかがえる。
- “X / and Y” や “X / or Y” という並列構造の接続詞の直前
接続している対象 X, Y に依存するが、“S and S” というように文同士の接続では、ほぼ確実にスラッシュを入れる。また、X, Y の長さに応じて、例えば “Japan and German” などであれば一気に読み進めスラッシュは入れない場合が多いが、かなり長い句である場合にはスラッシュを入れる。また、X, Y の統語構造に応じてその傾向はかなり異なる。
- “V NP / ADV / ...” という文中の副詞の直前直後

これについては被験者によってばらつきが大きく、常に成り立つわけではない。しかし、被験者によっては “next week” や “successfully” と副詞句 (または直前の V の目的語となる名詞句を) をセグメントとして独立させる傾向がある。

一方、スラッシュを入れる箇所が本質的に異なるものは、次のような場合である。

- “ADV / S” や “CONJ / S” といった文頭の接続詞・副詞の直後
被験者によってスラッシュを入れるかどうかは異なり、常に成り立つものではない。本手法では文頭の最初の1語は制約1より決してスラッシュが入れられることはない。しかし、文頭の接続詞は多くの場合読者に以後の話題の転換を示す。依存構造では接続詞は主文の動詞に係ることになるが、この依存関係は名詞句と動詞句の依存関係とはかなり趣が異なり、依存構造そのものに再検討の余地がある。
- “care for / NP” や “call on / NP” といった、いわゆるパターン
パターンとして部分構造を介さず直接チャンクを構成するような表現がある。その場合は現在仮定している依存構造では十分に対応ができない。詳しくは次節で議論する。
- 語長の影響
例えば、文頭からの主語を構成する句が同じ語数で、なおかつ同じ被験者であっても、物理的な語の大きさ、つまり語の長さが大きい語が連続する場合とそうでない場合にはスラッシュの入れる位置が前後する場合がある。語長が大きければセグメント長が短くなる傾向がある。
- 語の親密度の影響
2とも関連する事項であるが、語の親密度 (語に

対するなじみの程度)に応じてスラッシュの入れる位置が前後する場合がある。つまり、親密度が低い語については、理解に負荷がかかり、それを含むセグメント長は短くなる傾向がある。

4.3 議論

実験結果により得られた知見を踏まえ、本研究の今後の課題と方針について述べる。

統語範疇

被験者のスラッシュの入れ方は多様であったが、ほぼ唯一共通したのは補文の直前である。「先行詞+関係節」のような場合は構成される句の大きさに依存したが、「he said *S*」といった補文構造はほぼスラッシュが挿入された。今回の手法では現時点での統語解析の精度から、統語範疇は考慮することはなかったものの、ある特定の統語構造、補文構造や並列構造といった比較的特徴的な統語範疇のみを発見することは大きく精度を落とさないと予想される。上記のような傾向に限れば、ある程度本手法に加えていっても自動化の際の大きな問題とはならない。

また、我々は現在、同じ目的で確率モデルに基づいたスラッシュの挿入手法の開発を進めている [13, 10]。これは、均質な少量のスラッシュ付き英文をサンプルとするもので、どの統語範疇と統語範疇の境界、および長さにおいてスラッシュを入れる傾向があるか、ということを効率的に学習するものである。大きな確率でスラッシュが入る統語範疇間の境界が発見されれば、それらの傾向を本手法に加味していく予定である。

パターン

“have something to do with *A*”は「*A*と関係がある」、 “prefer *A* to *B*”は「*B*よりも*A*が好きだ」といったパターンは、英語の効率的運用に重要な役割を果たす [7]。これらは部分構造を介さず直接意味を取ることができるので、語彙チャンクや unanalyzed chunk とよばれる [9]。本稿の枠組みでいえば、パターンごとに依存構造が予め準備されていると考えることになる。つまり、パターンに対する依存構造を準備すればチャンク数の推移等も一意に決定される。現在連続なパターンについては取り扱えるよう実装を進めている。

問題となるのはどのような表現をパターンとして考えるか、ということであるが、中高英語参考書にある

パターンの重要度を振り所にするなど、実際に本システムを利用する場を想定しつつ検討を進めなければならない。さらにセグメント長を数える際にパターン全体(または不連続部分)を1語とするのか、といった問題が挙げられる。

セグメント長

セグメント長については、まず数え方、そして上限 θ の設定法の2つの問題がある。現在、セグメント長は単純にセグメント内の語数としているが、習熟していればいるほど実際にはさきのパターンのようにある程度まとまった部分表現を一括してチャンクとしていえると考えられる。また、実験結果でも述べたように、語の物理的な大きさや語の親密度も大きく関連していることが分かる。今後心理実験などでセグメント長の数え方を検討していく予定である。

また、セグメント長の上限 θ を決めるための方法も開発する必要がある。今回のような決め方が最適とは言えないが、仮に今回のような決め方にするとしてもプロフィール用の英文等がどのような性質をもつべきか、などの議論、検討が必要である。

5 まとめ

本稿では、スラッシュ・リーディング用教材の自動生成を念頭に、英文とその依存構造からスラッシュを入れる手法を提案した。単純な枠組みではあるが、簡単な心理実験では比較的良好な結果を得、さらに今後考慮すべき要因を洗い出すことができた。

今後の課題としては4.3節で述べた他に、実際に英語学習者を対象として本手法でスラッシュ付けした英文を教材としたスラッシュ・リーディングを実践し、学習効果を測り、その有効性を検討することも挙げられる。

謝辞

本研究について心理学見地から議論して頂いた福岡大学 二瀬由理先生に感謝致します。本実験の評価を手伝っていただきました九州大学情報基盤センター 原由加里女史、システム情報科学府自然言語処理研究室の諸氏、金内みつ子秘書に感謝致します。また同研究室 薬師寺亮太氏には有益な議論を頂きました。

参考文献

- [1] 日本電子化辞書研究所：EDR 英語コーパス ECO-V016E.
- [2] Ellis, N.C.: Memory for language, in P. Robinson (ed.), *Cognition and Second Language Instruction*, Cambridge University Press (2001).
- [3] 稲盛洋輔：英語の句読法辞典，インターワーク出版 (2003).
- [4] 西村友美：通訳技法「スラッシュ・リーディング」の英語教育への応用について (1)，京都橘女子大学外国語教育研究センター紀要，第1号 (1993).
- [5] Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, Vol.63, pp.81-97 (1956).
- [6] 門田修平編：英語リーディングの認知メカニズム，くろしお出版 (2001).
- [7] 門田修平：英語の書きことばと話しことばはいかに関係しているか，くろしお出版 (2002).
- [8] 門田修平編：英語のメンタルレキシコン，松柏社 (2003).
- [9] Nation, I.S.P.: Chunking and Collocation, *Learning Vocabulary in Another Language*, pp.317-343 Cambridge University Press (2001).
- [10] 高山英士：ALL IN ONE *Second Edition*，リンケージ倶楽部 (2003).
- [11] 田中省作，山本祥平，薬師寺亮太，富浦洋一：読解過程におけるナイーヴな Chunk の計算手法とマジカルナンバー 7±2，電気関係学会九州支部大会 (2003).
- [12] 寺島美紀子：英語「直読直解」への挑戦，あすなろ社 (2002).
- [13] 薬師寺亮太，富浦洋一，田中省作：スラッシュ・リーディング用英文テキストの生成—確率モデルに基づく英文へのスラッシュの挿入—，電気関係学会九州支部大会 (2003).