

Web を活用したドイツ語多読支援環境の構築

田中 省作 田畑 義之

九州大学 情報基盤センター 研究部

〒 812-8581 福岡市東区箱崎 6-10-1

E-mail: {sho@cc, tabata@rc}.kyushu-u.ac.jp

多読は、外国語学習において非常に有効な学習法の一つである。しかし、大学におけるドイツ語やフランス語といった初習外国語では、学習者に基本的な語彙や文法が定着していないことや、多読教材が十分ではないことから、直ちに多読を実践することが難しい。そこで、我々はドイツ語の多読の教材として、実文書に対して読解の障害となる表現に対して註釈を付与した HTML 文書を考える。また、この文書をドイツ語教師が作成するには、HTML に関する知識が必要となる、註釈付与の作業量が負担となる、といった問題があった。本稿では、多読教材としての註釈付き文書と、これらの文書の作成支援システムについて報告する。

Construction of the Environment for Extensive Reading of German Texts Using Web

Shosaku Tanaka and Yoshiyuki Tabata

Research Division, Computing and Communications Center, Kyushu University

Higashi-ku, Fukuoka, 812-8581 JAPAN

E-mail: {sho@cc, tabata@rc}.kyushu-u.ac.jp

Extensive reading is one of the effective strategies for learning foreign languages. However, for Japanese university students it is difficult to do this reading in a second language such as German, French and so on except English, because they don't have sufficient knowledge of their basic lexicons and grammar. Besides there are not so many materials for extensive reading of those languages. Then in German language classes we use HTML authentic texts with interlinear glosses for the expressions that may be difficult to understand in reading. In making these texts, there are some problems as follows: teachers should be required to be well aware of HTML. And loads of work of giving those glosses can be heavy. This paper reports the HTML texts with glosses as materials for extensive reading and the system for preparing those texts.

1 はじめに

多読 (extensive reading) は、外国語学習において非常に有効な学習法の一つである。多読の実践には、学習者が語彙や文法に関する基礎知識が備わっていることが必要とされる。そうでなければ、絶えず辞書や文法書を参照することになり、多読は難しい。大学における英語を考えてみると、学習者である学生は既にある程度の基礎知識を備えていることが期待できる。さらに、学習者の習得レベルに応

じた多読教材も潤沢である。よって、近年、英語の多読が大学の講義などにも積極的に導入されている。一方、大学における英語以外のドイツ語やフランス語といった外国語 (以後、初習外国語とよぶ) では、ほとんどの学生はさきの基礎知識が不十分であることや、多読教材が英語のそれに比べて極めて貧弱であることから、多読を直ちに実践することは難しい。

本稿では、大学でのドイツ語の多読を念頭に、初習外国語における多読教材について検討する。ドイ

ツ語の多読教材として、実文書に学習者がスムーズに読み進められるよう註釈を埋め込んだ註釈付き文書を示す。註釈は、主に形態素レベル(単語の意味、活用、品詞など)の情報である。次に、註釈付き文書の作成支援システムについて述べる。最後に、関連アプリケーションとの比較検討を行う。

2 初習外国語としてのドイツ語と多読

2.1 多読教材の要件

多読は、大量の文書を細かい内容にとらわれず、要点を掴み内容を楽しむ読み方である。近年、この多読が外国語の習得に効率的であることが報告されている[1, 2, 3, 14]。

文献[1]では、多読の実践の要件として10項目を挙げている。ここでは、多読教材に関する観点からそれらの要件を捉え直し、次の2項目に集約して述べる。

● 教材の量

多読では「学習者主導」であることが重要である。これは、多読教材の中から学習者自身が自分の興味で教材を選択し、読み進める。内容が面白くない、理解が難しい、といった場合には、途中で読解を中断し、教材を変えても構わない、というものである。そのためには、教師は広い分野にわたって学習者が関心をもつ教材を、大量に整備しなければならない。

● 教材の難しさ

教材の難しさが学習者の習得レベル内にあることも重要である。未知語が多かったり、品詞や文法役割が同定できず、絶えず辞書や文法書を見ながら読み進めることは、本質的に多読とはいえない。教師は学習者の習得レベルに応じた教材を、整備しなければならない。

そこで、英語では、多読の最初の手がかりとして、段階別読本(抄訳本: graded reader)を多読教材に用いることが多い。段階別読本は、習得レベル別に使用語彙や構文を制限し、原文を意図的に書き直したものである。英語の段階別読本は、小説や時事問題など分野も広く、量的にも充実しているといえ

る¹。だが、段階別読本そのものも、一般的には次のような問題が指摘されている。

● 内容の新しさ

段階別読本を作成するまでに時間を要し、新しい内容のものが少ない。

● 表現の自然さ

段階別文書では恣意的に表現を簡約化するため、低頻度語や、普通に使われるような統語的・意味的用法が削除される場合がある。その結果、不自然、単調な表現になりやすい。

英語については、近年、これらの状況はかなり改善されつつあるが、初習外国語の段階別読本では、これらの問題はまだ深刻で、量も英語のそれと比べると極めて貧弱である。現時点では、初習外国語の段階別読本による多読は十分とはいえない。

そこで、我々はドイツ語の多読教材として実文書(authentic text)を用いることにする。実文書は母語話者が読むことを念頭に書かれており、量はもちろん、さきに述べた段階別読本における問題も回避される。問題となるのは、実文書と学習者の習得レベルが合致しない場合である。そこで、実文書に対し学習者が読み進めるのに障害となりそうな表現に対して註釈を与えることで、多読教材として活用する。

2.2 ドイツ語における「読解」の障害

大学における初習外国語では、ほとんどの学生は英語の学習経験が既にあるので、外国語自体には抵抗が少なく、一般的な文法現象などは、英語と対比することによって、比較的容易に理解できることが多い。ドイツ語に限定していえば、言語的に英語との親和性が高いこともあり、文中の単語の意味や品詞²の同定、つまり形態素レベルの理解ができていた状況下では、大意を把握する程度の大雑把な統語レベルの理解³は、さほど難しくない。

¹読本という形態ではなく、英語学習者向けのための英字新聞や雑誌といった形でも多く提供されている。

²例えば、“die”は定冠詞か、関係代名詞なのか、という意味での曖昧さがあるが、多くの場合、単語の接続関係などの形態素レベルで同定される。仮に“die”が関係代名詞であるとき、先行詞がなにか、つまり単語(句)間の対応関係は、統語レベルの問題である。

³句構造の同定、すなわち関係代名詞と先行詞、主語・述語の対応付けなどが挙げられる。

一方、ドイツ語に触れた時間が絶対的に少ないため、基本的な語彙の不足、品詞の同定に不慣れである、といった問題がある。また、既知の単語であるにもかかわらず活用が原因で理解できない場合もある。つまり、学生が文理解に失敗する場合は、形態素レベル(単語の意味、活用した単語の基本形、品詞の同定)での失敗に起因することが多い、ということが予想される。そこで、註釈の内容は、形態素レベルの情報を中心に、学生の習得レベルに応じて、難しいと思われる単語や成句といった表現に対して与えていくことにする。

また、外国語に限らず読解では、文書中で明示される情報の他に、背景知識が重要な役割を果たす。例えば、「Gehard Schröder」に対して固有名ではなく「現ドイツ首相」という情報が重要である場合もある。註釈としては、これらの背景知識も随時与える。

3 Webを活用した多読支援環境

3.1 註釈付き文書

ドイツ語の実文書に対して註釈を埋め込んだ註釈付き文書について述べる。2.2節で述べたように註釈には、その表現の形態素レベルの情報を中心に、背景知識に関する情報なども与えられる。

註釈付き文書は、HTMLで記述され、註釈もHTMLのソースの一部として埋め込まれる。学習者はWebブラウザさえ使用できる環境にあれば読むことができるし、文書がローカルに置かれていれば、ネットワークに繋がっている必要もない。Webブラウザでこの文書を開くと、註釈が埋め込まれている表現は色付きになっており、その表現の上にマウスを合わせると、サブウィンドウが開いて読解情報が表示される(図1)。

この註釈付き文書の主な特徴は次の3点である。

- 註釈で意味を記述する場合は、単なる辞書項目の羅列ではなく、その文脈下での意味を端的に表したものである。本来、読解の重要な目的の一つは、多義な表現の適切な意味を文脈から推測する力を養うことである。しかし、我々は、多読の場合は、むしろ読みが中断されることなく、スムーズに読み進むことの方が重要であると考えた。

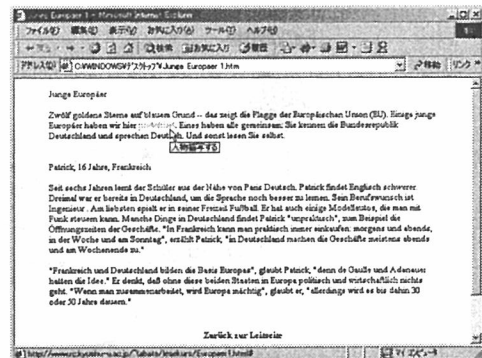


図1: HTMLの註釈付き文書(“porträtiert”に対して「人物描写する」という註釈が表示されている)

- 註釈は常時提示されるのではなく、学習者自身がマウスを表現に合わせるという能動的なアクションがあつて初めて提示される。よつて、註釈ばかりに頼つて読み進める、といったことは抑制される。
- 註釈が付与されている表現の直下に表示されるので、大きな視線の移動が無い。

次に、その作成過程について考えてみる。ドイツ語の実文書を基とするので、広い分野にわたつて、しかも新しい内容のものが容易に入手可能である。しかし、註釈を埋め込む表現、註釈の内容そのものは、学習者の習得レベルや文脈に依存するため、予め教師がチェックしなければならない。よつて、註釈付き文書を人手で大量に整備するには、大変な労力がかかる。さらに、HTMLや文字コードを考慮しなければならない、といった問題もある。そこで、これらの作業を軽減するための註釈付き文書の作成支援システムを構築した。

3.2 註釈付き文書の作成支援システム

ドイツ語教師が註釈付き文書を一から作成するには、主に次のような問題がある。

1. HTMLや文字コードに関する知識が必要となる。
2. 註釈の埋め込み作業が大きな負担となる。

1. は、システムで完全に解決される。2. については、註釈は形態素レベルの情報が主であるので、自然言

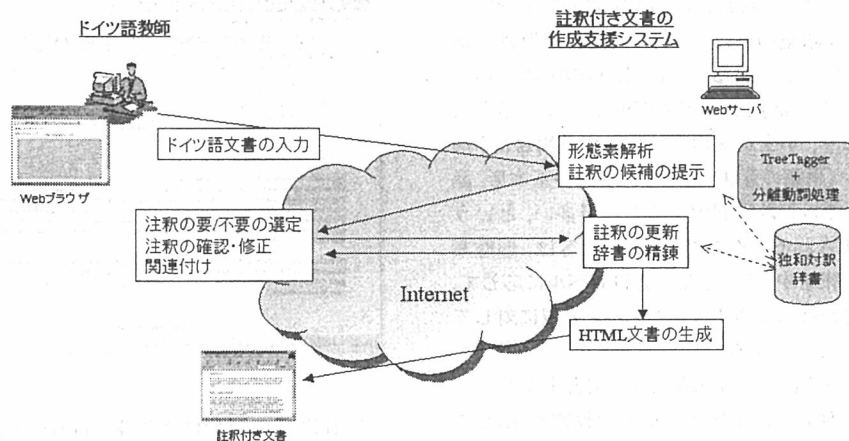


図 2: 註釈付き文書の作成の流れ

語処理技術の要素技術である形態素解析が有用である。形態素解析は、文を入力し、文中の単語とその品詞の同定を行うものである。この技術を利用して、注釈の候補を予め教師に提示することが可能となる。

註釈付き文書の作成支援システム（以後、本システムとよぶ）は、Webサーバとして構築されており、ネットワーク経由でWebブラウザの一般的な操作で使うことができる（図2）。本システムによる註釈付き文書作成の流れは次のようになる。

Step 1. ドイツ語文書を入力する（図3）。

Step 2. 入力されたドイツ語文書を形態素解析し、システムの独和対訳辞書（3.3節）を基に注釈の候補を提示する。

Step 3. Step 2.の結果を教師に提示し、注釈のチェックを行う（図4）。

Step 4. 註釈付き文書を生成する。

各Stepについて説明していく。

Step 1. ドイツ語文書の入力

本システムのWebページを開き、図3のフォーム部分にドイツ語文書を入力する。キーボードから入力してもよい⁴、ドイツ語のWebページからコ

⁴ドイツ語固有のアルファベットであるウムラウトは、Äを#A, ßを#sなどで代替している。

ピー&ペーストしても構わない。ここでの文書のレイアウトは、最終的に生成される註釈付き文書にも反映される。

Step 2. 形態素解析と注釈の候補の提示

Step 1.で入力されたドイツ語文書を形態素解析し、単語毎に独和対訳辞書から注釈の候補を提示する。この独和対訳辞書の詳細については、次節で述べる（3.3節）。ドイツ語文の形態素解析にはTreeTaggerを使用した⁵。

なお、TreeTaggerは後述する分離動詞に対して、分離前綴りの同定は行いが、基本動詞部分との対応関係までは求めない⁶。分離前綴りと基本動詞部分の解析部分については、自前で作成した。

Step 3. 教師のチェック

Step 2.で各単語毎に付与された注釈を教師に提示し、教師は、次のような操作を行う。

- 注釈を埋め込む単語を選択する。品詞単位で一括して埋め込む/埋め込まない、を指定することもできる。

⁵<http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/DecisionTreeTagger.html>

⁶分離動詞は、前置詞（分離前綴り）と動詞（基本動詞部分）からなる動詞である。文中では分離前綴りと基本動詞部分が分離して出現し、その対応を求めるには、形態素レベルの情報のみでは不十分で、統語レベルの情報も必要となる。

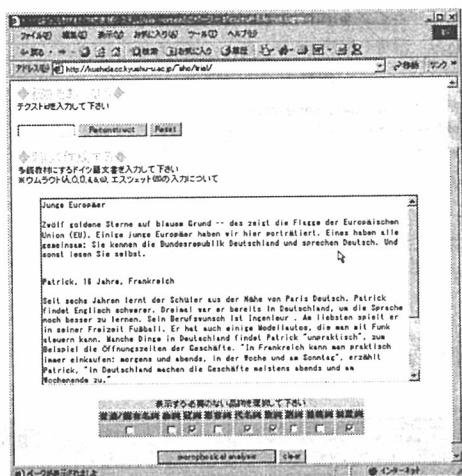


図 3: ドイツ語文書の入力

- 提示された註釈の候補に修正を加える。
- 修正した註釈を、システムの独和对訳辞書に反映させる。
- 成句など、単語間の関連付けを行う⁷。

Step 3. での註釈の修正情報は、システムの独和对訳辞書の精練に使用され、更新操作を行うことで、直ちに作成中の文書の註釈にも反映される。

図 5 は、形容詞 “jung” に対して、註釈の埋め込み (ADD 欄)、註釈に対する修正を辞書に反映 (ENTRY 欄)、註釈の候補から「若い」を選択した例である。

Step 4. 註釈付き文書の生成

註釈付き文書ではドイツ語と日本語が混在するため、Unicode (UTF-8) の HTML 文書を生成する。

現状の独和对訳辞書では、登録語数が不十分のため、註釈の候補が提示される内容語の割合は、概ね 5~6 割程度である。教師が新たに与えた註釈は、随時独和对訳辞書に加えられるので、将来的に、独和对訳辞書が十分整備されれば、教師の実作業は註釈

⁷図 4 の CONJ 欄に形態素番号を入力し、関連付けを行う。関連付けられた単語のうちの一つにマウスを当てられると、関連付けられた他の単語と共にハイライトされる。分離動詞については、分離前綴りと基本動詞部分が自動的に関連付けが行われる (図 10)。

NO	NOTE	POS	DIR	CONJ	ADD	ENTRY	SUPPORT INFORMATION
0	Junge	ADJA	jung	☐	☐	☐	新しい, 若い
1	Europäer	NN	Europäer	☐	☐	☐	欧州オープン
3	goldene	ADJA	golden	☐	☐	☐	金色
4	Stemme	NN	Stemm	☐	☐	☐	鼻
6	blauen	ADJA	blau	☐	☐	☐	青い
7	Grund	NN	Grund	☐	☐	☐	地面, 土地, 階, 土台, 基礎, 理由 (Zusatz: 根拠, 動機, 原因, 都合)
9	das	POS	d	☐	☐	☐	関係代名詞
10	zeigt	VFIN	zeigen	☐	☐	☐	見せる, 提示, 指示, 示す, 発表, 浮かべ
12	Flage	NN	Flage	☐	☐	☐	旗, 面旗

図 4: 教師による註釈の確認・修正

から不適切、冗長な情報を削除するだけでよいことになる。

3.3 システムの独和对訳辞書

本システムで作成される註釈付き文書は、元の文書の著作権が許せばそのまま本システムにストアし、共有、公開したいと考えている。そのため、本システムの独和对訳辞書は、市販のものではなく独自に整備したものを使用している。これらの辞書は一から作成するのはなく、既存の電子化資源、特に英語をピボットとしてフリーの独英辞書と英和辞書からの合成を試みている [9, 10, 12]。

辞書項目は、ドイツ語表現 (g) に対して品詞 (p) と、日本語訳 (j_i)、過去の註釈の埋め込み作業でその j_i が選定された頻度 (f_i) の組の集合である ($1 \leq i \leq m$)。

$$g : p : \{ \langle j_1, f_1 \rangle, \dots, \langle j_m, f_m \rangle \}$$

発音情報や語源情報といったものは多読の際には不要なので含めない。

Step 2. の g の註釈の候補の提示では、選定頻度に従い降順で日本語訳を並べる。このとき、教師によって選択された j_i の選定頻度 f_i は 1 インクリメントし、より頻繁に選択されるものが、先に提示されるようにする。

また、独和对訳辞書は、教師の註釈に対する修正によって動的に拡充される。特に通常の辞書では弱い固有名詞や新語、時事用語に強い辞書を構築でき

形態素 番号	表記	品詞	原形	関連付け	注釈埋め込み	注釈修正を 辞書へ反映	注釈	
NO	NOTE	POS	ORG	CONJ	ADD	PREF	ENTRY	SUPPORT INFORMATION
Junge Europäer Zwiif goldene Sterne auf blauem Grund... die zeigt die Flagge der Europäischen Union (EU).								
0	Junge	ADJA	jung	☐	☒	☐	☐	新しい, 若い
↓								
0	Junge	ADJA	jung	☐	☒	☐	☒	若い

図 5: 注釈の修正の例

と思われる。例えば, “Gehard Schroder” に対して「ゲアハルト・シュレーダー (現ドイツ首相)」といった通常の辞書には記載されないような説明的な情報も時事に合わせて記載される。

現在, 多読用の独対訳辞書の登録語数は 37,881 語である。

4 関連アプリケーションとの比較

ここでは, まずドイツ語以外の外国語対象のアプリケーションも含め Web を利用した読解支援アプリケーションとの差異を述べる。次に, 現在実際に使用可能なドイツ語対象の読解支援アプリケーションとの比較を行う。

4.1 Web を活用した読解支援アプリケーション

Web を活用した読解支援のアプリケーションとしては, ドイツ語を対象としたものではないが, 朝日新聞の「Popup 英和ニュース」⁸(図 6) や「日本語読解学習支援システム リーディング チュウ太」⁹(図 7) などがある。

前者は, 学習者が下線付きの英単語や成句をクリックすると, 辞書項目をサブウィンドウに提示する。内容も新しく興味深い記事が数多く提供されている。

後者は, 学習者が入力した日本語文を形態素解析し, 内容語に対する辞書項目を別フレームに表示する。学習者自身が関心の高い任意の文書を入力する



図 6: Popup 英和ニュース

ことができる。

いずれも, 特に後者は学習者の多様な興味に対応でき, また十分な情報を提示することで, 読解の支援を行う。しかしながら, 辞書項目をそのまま提示するため, 意味を多数有す基本単語などの場合, 学習者に提示される情報が膨大となり, 多読という観点では必ずしも適当ではない。

本稿の注釈付き文書では, 教師のチェックを一度受けるので, 注釈がその文脈下で端的で適切なものであることが保証される。適宜, 説明的な情報も記述される。その一方で, 文書の作成を完全に自動化できず, 教師自身にチェックという実作業が伴うことになる。

⁸<http://www.asahi.com/english/popup/>

⁹<http://language.tiu.ac.jp/>

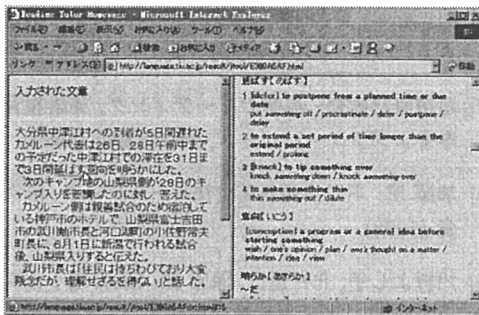


図 7: リーディング チュウ太

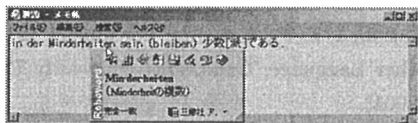


図 8: “Minderheiten” に対する Roboword の表示

4.2 ドイツ語を対象としたアプリケーション

現在、日本語をベースとしたドイツ語学習に関するアプリケーションは、あまり充実していない。ドイツ語の多読に特化したものではなく、読解支援といっても主に単語の辞書引きの自動化に重点が置かれたものである。ここでは、ドイツ語を対象とした読解支援の代表的なアプリケーションとしてテクノクラフト社¹⁰の Roboword を例に挙げる。

Roboword は、マウスが指す単語の辞書項目を、サブウィンドウに表示するアプリケーションである。表示される情報は、辞書項目の羅列であり、読みには不要の発音情報、語源情報といったものまで表示する。Roboword の辞書は、通常の紙の辞書を単純に電子化したもので、それによる弊害もある。例えば、“Minderheiten” に対しては「Minderheit の複数形」と表示されるだけである (図 8)。

また、Roboword はドイツ語文を形態素解析をせずにパターン・マッチングで単語の品詞同定を行うため¹¹、分離動詞などには全く対応できない。例えば、“an|kommen”(到着する) という分離動詞¹²は、

¹⁰<http://www.technocraft.co.jp/>

¹¹ドイツ語は、分かち書きされており単語分割は容易、しかも名詞は大文字で書き出すといった書法があり、品詞同定はパターン・マッチングだけでもある程度の精度が得られる。

¹²分離動詞を記述する場合、分離前綴りと基本動詞部分が分

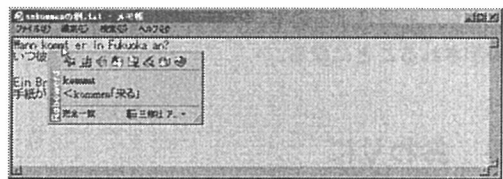


図 9: 分離動詞の基本動詞部分 (“an|kommen” の “kommt”) に対する Roboword の表示



図 10: 分離動詞の表示 (“an|kommen” の “kommt”) に対する註釈付き文書の表示)

分離前綴り “an” と基本動詞部分 “kommen” が、次のように現われる。

Sie kommt morgen in Berlin an.
(彼女は明日ベルリンに 到着する。)

単純にパターンだけを参照したのでは、“kommen” がそれだけで動詞なのか、分離動詞の基本動詞部分なのか、“an” が前置詞なのか、分離前綴りなのか、ということが決定できない。その結果、Roboword は “kommt” に対しては、動詞 “kommen” の辞書項目、“an” に対しては前置詞 “an” の辞書項目という具合に、誤った情報を提示することとなる (図 9)。一方、本システムでは、形態素解析を施し、分離前綴りと基本動詞部分の対応関係まで求めているので、システムが与える註釈は、基本動詞部分の “kommt” に対しては「an|kommen: 到着する」、分離前綴りの “an” に対しては「ankommen の分離前綴り」となる (図 10)¹³。

分離動詞以外にも、“weiß” が¹⁴、形容詞 “weiß” (白い) なのか、動詞 “wissen” (知っている) の活用したものなのか、といった場合も Roboword では十分に対応できない。本システムでは、形態素解析で品詞

かるように “|” を挿入する。

¹³分離前綴りと基本動詞部分は自動的に関連付けが行なわれるので、分離前綴り (基本動詞部分) にマウスが合えば、基本動詞部分 (分離前綴り) と共にハイライトし、どのように対応しているかも分かる。

をきちんと同定するので、この場合も正しい情報が提示されることになる。

5 おわりに

本稿では、多読教材としての註釈付き文書と、その作成支援システムについて述べた。ここではドイツ語を対象としたが、ドイツ語固有の処理を置き換えれば、他言語でも、ほぼ同様のアプローチが可能である。

今後の課題としては、この多読環境の評価が挙げられる。評価は、(i) 註釈付き文書の学習効果に対する評価、(ii) 作成支援システムの作業効率に対する評価に大別される。(i)については、小規模ではあるが、註釈付き文書を教材とした多読演習を実践しており、[11, 13]で報告している。多読による付随的学習の効果(例えば、語彙力の伸び)なども評価していきたい。(ii)については、HTMLや文字コードに関する作業量は確実に軽減される。註釈の埋め込み作業をどう定量的に評価するか検討していく。

本来、多読教材は学習者が各自の興味に従って選択できるのが望ましい。作成された註釈付き文書については、著作権に問題がないもの¹⁴については共有し、学習者自らが自由にそれらを読み進められるような環境にしていきたい。

なお、註釈付き文書の作成支援システムについては、試用版を、

<http://lengua.cc.kyushu-u.ac.jp/german/er/>
で公開しています。

謝辞

ドイツ語文の形態素解析に TreeTagger を使用させて頂きました。Stuttgart 大学の Helmut Schmid 博士に感謝致します。

参考文献

[1] Day, R. R. and Bamford, J.: Extensive reading in the second language classroom, Cambridge University Press (1998).

[2] 門田修平, 野呂忠司: 英語リーディングの認知メカニズム, くろしお出版 (2001).

[3] 垣田直巳監修: 英語のリーディング, 英語教育学モノグラフ・シリーズ, 大修館書店 (1984).

[4] 加藤 宏: ドイツ語新正書法ルールブック, 郁文堂 (1998).

[5] 川村よし子, 北村達也, 保原 麗: EDR 電子化辞書を活用した日本語教育用辞書ツールの開発, 日本教育工学雑誌, 第 24 号, pp. 7-12(2000).

[6] 講談社: ルビー・ブックス.

[7] Nation, I. S. P.: Learning Vocabulary in Another Language, Cambridge University Press (2001).

[8] Schiller, A., Teufel, S. and Stöckert, C: Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS, Seminar für Sprachwissenschaft (1995).

[9] 白井 諭, 山本和英, 白 京姫: 対訳辞書作成のための英訳辞書の照合, 電子情報通信学会 信学技報 NLC2000-71, pp. 17-23 (2001).

[10] Tanaka, K. and Umemura, K.: Construction of a Bilingual Dictionary Intermediated by a Third Language, *Proc. 15th International Conference on Computational Linguistics*, pp. 285-292 (1994).

[11] 田中省作, 田畑義之: Web を活用したドイツ語の多読支援環境とその構築, 信学技報 教育工学 ET2002-20, pp. 55-60 (2002).

[12] 田中省作, 富浦洋一: 類語集合による英語を介して導出した対訳候補の絞り込み, 第 1 回情報科学技術フォーラム FIT2002 (accepted).

[13] 田中俊明, 田畑義之: マルチメディア時代のドイツ語教育, 九州大学出版会 (2000).

[14] 津田塾大学言語文化研究所編: 英文読解のプロセスと指導, 英語教育 21 世紀叢書 大修館書店 (2002).

¹⁴例えば, Project Gutenberg(<http://promo.net/pg/>) で収集された文書を基としたもの。