

誤用データ機能を備える WEB 中国語作文 添削支援システム設計と開発

(早稲田大学政治経済学術院・中国語遠隔教育) 砂岡和子 ksunaoka@waseda.jp
(早稲田大学国際情報通信研究科院生・情報通信工学) 劉松 liusong@suou.waseda.jp

要旨:

早稲田大学が実施する少人数ライブ型口語レッスン「チュートリアル中国語」プログラムに併設する、Web 中国語作文添削支援システム「写作諸葛」の設計コンセプトと添削実績について、3年間の運用実績に基づいて発表する。

「写作諸葛」はオンラインでの字句添削に加え、統一文法添削マークを併用することにより、従来の添削者間の評価基準の不統一、添削記述方式の不統一を改善すると同時に、回収した作文添削データを誤用コーパスとして蓄積し、専用統計エディターで誤用分析を行う。誤用統計は個人およびレベル別誤用傾向を把握する資料となり、次回の出題や学生指導に生かされる。発表では、オンラインエディターの設計理念と、日本人中国語学習者用の誤用コーパスの開発、統一誤用マークタグによる作文添削、および添削文のデータベース化と、統計分析に基づいて改良した自動添削の新しい枠組みについて述べる。

キーワード: Web 中国語作文 協調学習 誤用コーパス

1 遠隔語学レッスン用 Web 作文・添削

「写作諸葛」は早稲田大学が2003年から開設するライブ型口語レッスン「チュートリアル中国語」コースの履修生を対象に、科目専用の学習サイトで展開する Web 作文・添削システムである。「チュートリアル中国語」は英語やその他外国語チュートリアル同様、主に学部生を対象にしたコミュニケーション重視の口語教育プログラムである。学習者4名対ネイティブインストラクター1人による対面型中国語会話レッスン (Fig. 1) と、台湾師範大学や中国北京大学の教員による遠隔テレビ会議方式レッスンを組み合わせたハイブリッド授業 (Fig. 2) の2タイプがある。



授業用プラットフォームとして開発した「Tutorial Chinese Platform (以後、TCPと略称)」は静的素材がメインのE-Learningとは異なり、遠隔方式を採用したリアルタイムの対話型口語教学に対応できるダイナミックな設計とした。ネットワークの双方向性を生かし、TCPは教場での情報を直ちに学習者へフィード・バックし、国内外の教員と学習者の緊密なユーザー連合を組織する。現在までにレベルわけWebテスト、テキスト準拠CALLドリル、作文添削、学習履歴の記録、ならびにオンラインレポート、ウェブアンケート、BBSなどサイトを運用している^[1]。

作文・添削サイトは、口語会話でもとすれば疎かになりがちな文章表現を補完するため設置され、2003年10月システム運用後2005年3月31日まで、単位登録したユーザー数は学習者計366名、添削者延べ14名、教員延べ11名、添削文計1228

篇、誤用マーク付きフレーズ総計17866余条を回収し、範疇別に分類した誤用コーパスが蓄積されている。

本稿ではTCPのWeb中国語作文・添削支援システムの特徴を紹介し、過去一年半の添削データの蓄積、ならびに誤用コーパスに基づく自動作文添削システム開発の展望を述べる。

2. システム設計コンセプト

本添削システムの特徴を作文投稿と添削の流れに沿って、以下に挙げる。

(1) **学習課題の適正管理**: 作文課題は毎回の口語レッスンの学習内容と連動したテーマについて、レッスン終了直後、それぞれ使用すべき重要表現と字数を指定し、初級用と中上級用の二題が出題される。外国語学習者に完全な自由作文を課すと、誤用が多出し添削基準の統一が難しい。個々の語学レベルに適した学習済みのテーマについて、キーワードが指定された短い半課題文が添削効果を生む。

(2) **テーマ選択による学習意欲喚起**: 北京と台湾にいる教員が同じテーマについてそれぞれ初級用、中上級用の作文課題を出すことで、受講者に難度と文化差の選択権を与え、学習意欲を喚起する。またプログラム上で作文課題提示時期と提出期限をレッスン終了後一定期間に限定することで、学習者の緊張感を維持する教育的効果が期待できる。(Fig. 3参照)

(3) **スムーズな作文・添削オンラインエディター**: Web上で違和感なく投稿・添削できる編集機能をもつ専用のRich textベースのテンプレートを開発。アンダーライン、取り消し線、太字、文字色の変更等、ワードテキストエディターに近い感覚で作文を投稿できる。添削者も、挿入、訂正、削除等、ペーパー添削と同じ要領でウェブブラウザを操作できる。(Fig. 4参照)

(4) **多言語対応インターフェイス**: 学習者と添削者はそれぞれ日本・北京・台湾に分散し、かつ母語の文字フォントが異なるため、「写作諸葛」は時空間の制限がないWebベース添削シ

システムを採用し、Unicode による多言語対応可能とした。ネットワークに接続できる環境であれば、通常のインターネット掲示板書き込みと同じ感覚でいつでもどの地域からでも作文練習と添削ができる。

(5) **受理・返却情報の明示**：作文提出と返却にタイムラグがあると学習効果を減じる。本作文システムは作文提出・受理・添削記録を瞬時に反映し、作業の流れが一目瞭然となるインターフェースを開発した。(Fig. 4)

(6) **誤用添削記号の統一**：作文エラーは単純なスペルミスや語彙の使い間違いなど単純な誤りから、語法や発想法など概念の誤用に至るまで多種多様であり、これらに対して一律に添削をおこなえば修正ポイントが拡散し、多大の労力にも拘わらず学習効果は薄い。「写作諸葛」は統計的に常見誤用例を下記の8種のタイプに分類し「誤用添削マーク」として標識化して添削を行う。

誤用類別タグ一覧

- L：論理性の欠如
- D：欠落
- E：余剰
- M：でたらめな造語
- P：語順の誤用
- C：組み合わせの不適
- R：指すものが不明
- W：用字や符号の誤用

統一マーク適用により学習者の誤用と添削者の修正情報に自動的にXML形式のタグ付与が可能となり、誤用データを自動的に蓄積できる。複数の添削者による評価のバラツキ縮小にも益する^[2]。(Fig. 4 参照)

(7) **優秀作文表彰システムによる見直し奨励**：添削結果の見直しを奨励するため、提出作文に評価点を付け、かつ優秀作文を表彰するシステムを設計した。添削済みの高得点作文は個人情報削除後、優秀作文リストに一覧表示され、全ユーザーが閲覧できる。他の投稿者の作品と比較することによって、自身の誤用に気づき、自立的な矯正を促すことが期待できる^[3]。(Fig. 3 参照)

(8) **誤用タグの自動蓄積と統計分析**：誤用タグを専用の検索・統計エディターを用いて、学習者全体あるいは個人の誤用履歴の分析を行う。結果は次の作文練習のポイント決定や自動誤用検出等、言語教育と学習者用コーパス構築に還元できる。(Fig. 7 参照)

(9) **BBSでの協調学習支援**：作文テーマごとにクラス単位、あるいは学習者全体で討論を行うスペースで、作文課題や質問を投稿。他の学習者や添削者との相互学習を支援する^[3]。討論記録はデータベースに格納され、後日参照可能である。(Fig. 6 参照)

优秀作业					
Text	Level	Title	Point	Edit	
第22课	初級	我最爱看的新闻类型	8	view	
第22课	初級	日本的媒体现状	8	view	
第22课	中高級	新闻媒体的社会责任	8	view	
第22课	初級	我最爱看的新闻类型	8	view	
第22课	中高級	新闻媒体的社会责任	8	view	

等待评分作业						
Student	Text	Level	Title	Teacher	Submit Time	Edit
	第19课	中高級	我了解的医疗保险制度		2005-02-10 20:19:39	view
	第22课	初級	日本的媒体现状		2005-02-04 17:41:53	view
	第21课	初級	我对战争的看法		2005-02-04 17:35:28	view
	第20课	初級	关于“年金”改革的看法		2005-02-04 17:25:19	view
	第19课	初級	我关心的政治问题		2005-02-04 17:19:58	view
	第19课	初級	我关心的政治问题		2005-01-24 15:34:12	view

已评分作业						
Student	Text	Level	Title	Remark by	Remark Time	Edit
	第22课	初級	我最爱看的新闻类型		2005-01-13 09:44:20	view
	第22课	初級	日本的媒体现状		2005-01-22 13:12:13	view
	第22课	中高級	新闻媒体的社会责任		2005-01-16 10:20:45	view
	第22课	初級	我最爱看的新闻类型		2005-01-13 22:29:47	view
	第22课	中高級	新闻媒体的可信度		2005-01-24 01:00:49	view

The screenshot shows a web-based writing interface. At the top, there are fields for 'Text' (第1课), 'Remark by', 'Level' (中高級), and 'Title' (自我介绍). The 'Description' field contains a prompt in Chinese. Below this, the 'Student' and 'Point' (7) fields are visible. The main 'Content' area shows a student's text in Chinese with various colored markers (blue, red, green) indicating errors. A toolbar with icons for bold, italic, underline, and other text formatting is positioned above the text area.

Fig. 4 作文編集・添削インターフェース

This screenshot shows a text editor window with a Japanese sentence: '我家养了三只猫，两只乌龟和很多的小鲤鱼。我们给这三只猫取了名字，金枪鱼和鳗鱼。我的妹妹喜欢吃生鱼，所以我们给他们这样的有意思的名字。三只乌龟和鲤鱼没有名字。他们都看见一样的。我小的时候，我家也养过两只小鸟。他们是托在学上的白文鸟的。他们很可爱，我想再养小鸟。' Below the text, several XML tags are shown, such as '<E>三只猫取了名字</E>', '<W>P</W>', and '的', which correspond to the error types defined in the text.

3. 誤用コーパスの設計と分析

「写作諸葛」は学習者の常見の誤用を統一文法添削マークを用いて添削することで、人手による評価のバラツキを最低限にとどめると同時に、データベースに誤用データを自動蓄積でき、専用エディターでの検索・統計を可能にした。(Fig. 5)の上

部が編集・添削モジュールでの作文インターフェース、下部が保存用のタグ付き形式のデータベースである。データベース全文のフレーズ単位に対し、誤用の種類、作文内容キーワード、作文題名と説明文、難易度、評価点数、及び提出者の個人情報等を、前方一致、後方一致、部分一致で検索できる。これらの検索手段によって学習履歴全体と個人別の作文提出状況、及び誤用頻度変化を統計でき、個々の誤用傾向が判り、自習や指導に役に立つ。誤用コーパス分析にも利便が多い。(Fig. 7 参照)

ITEM	EXAMPLE
性別	1--Male 2--Female
提出時間	LIKE 2004-01-20 =2004-01-20 =2004-01-20 >=2004-01-20
点数	=0 =8 <=0
難易度	1--Basic 2--Intermediate

1	誤り種類/Keyword	LIKE	[L]
2		LIKE	
3		LIKE	
4		LIKE	

検索式 例: E1*2791

AND(*) OR(+) NOT(-) ()

Search Reset Clear

“誤用種類はL”の条件での検索結果

ID	題目名	提出時間	難易度	点数	センテンス
15015	大家看了雅典奥运会的	2004-11-17 21:57:58	2	7	[L]我一定想在当地感受运动给我们的感动。
15019	据说长期使用手机有害	2004-11-17 23:03:33	2	7	日本社会已经变为人们没有手机[L]则不来。
15035	手机固然带给我们许多	2004-11-18 00:11:25	2	7	不能立刻通话的时候[L](不能立刻連絡到他的時候)他们不知道应该了。
15038	手机固然带给我们许多	2004-11-18 00:11:25	2	7	对有人好处[L][P]手机对人而言有好處也有壞處。
15136	當我們發生事故時,第	2004-11-18 12:28:08	2	8	可是工作是最重要的[L]雖然你們是社會上默默付出的一群人。

2003年10月から2005年11月の間に回収した添削フレーズを解析した結果、全体の1/3にあたる5308条のフレーズに誤用がみられ、中上級作文は初級と比べフレーズの平均長が長く、異なり単語数も増え、複雑な文になることがわかった。(Table 1)

	初級	中高級	全体
篇数	678	550	1228
フレーズ数	7748	9244	16992
字数	66554	95926	162480
フレーズの平均長	8.6	10.4	9.6
異なり単語数	4585	5594	7565

Table 1: レベル別語彙リストの全般

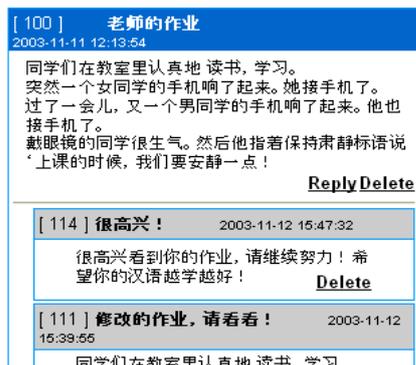


Fig. 6 BBS インターフェース

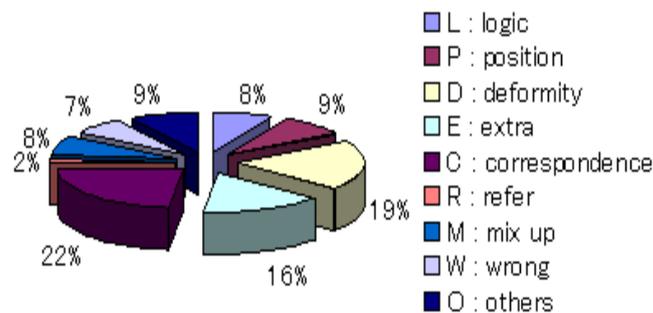


Fig 8: 誤用頻度分布

誤用の種類の分布から、「主語と述語、動詞と目的語などの、組み合わせの不適」と「述部、動詞アスペクト(態)などの欠落」、「接続詞や助詞、主語などの余剰」の3種の誤用がそれぞれ22%、19%、16%と上位を占め、残り5種の誤用頻度は低い結果が出た。これらの誤用分布の結果は、次章で述べる第二次「写作諸葛」の誤用タグ再定義の基礎データとなった。また出題テーマの選定をこれら統計的誤用コーパスに基づいて決定できる。(Fig. 8 参照)

4. 仮投稿プラットフォーム設定による自動添削へ向けて

現行添削システムの問題点は、誤用タグ付与の曖昧性に起因する添削結果の消化不良と、同類誤用の再出にある。解決策として従来の「文法添削マーク」を用いた添削方式を改め、作文表現失敗のタイプを周辺知識に広げ、かつ学習者同士が協調して正解知識を獲得する学習支援型プログラムの開発へと発想を転換して実験を行っている。

これは従来の個別投稿を、学習者同士が相互に協力して課題解決を行うプロセスへと転換し、プログラムに組み込むもので、正規投稿前の誤用コーパス参照機能と、仮投稿プラットフォームの設置による二段階添削方式を採用する。(Fig. 9 参照)

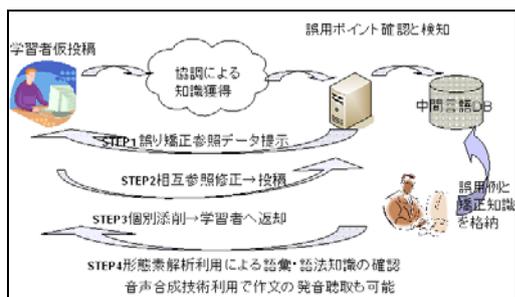


Fig 9: 協調的学習支援型添削プログラム

仮投稿プラットフォームに提出された学習者の作文は、いったん同テーマの作文課題を行っている学習者間で相互閲覧される。この間添削者はまだ修正を加えない。学習者同士が比較参照することで、かなりの誤用は修正されることが第一次「写作諸葛」添削プログラムのBBS活動で証明されている。学習者は誤用コーパスを参照しつつ、互いの作文を見比べることで自己修正を行う。修正を経て再度投稿された段階で、添削者がはじめて添削に参加する。すでに共通の誤用に対して自己修正が加わっている段階であるから、誤用範囲が限定されると期待できる。(Fig. 10 参照)

添削者の誤用タグを見直す必要から、添削コーパスから「旅行」に関するテーマの作文を約1万句抽出し再分析し、分析統計を行った結果、学習者の作文の失敗のタイプを以下のように再定義することにした。

学習者の作文の失敗のタイプ

- 1 学習者が自覚可能な失敗 ⇒ 個別知識確認タグ
例; 既習、テキスト化された学習項目
- 2 自覚しにくい失敗 ⇒ 周辺知識喚起タグ
例; 思い込み、不注意、非テキスト情報
- 3 本人は自覚できない失敗 ⇒ 既知識から連想タグ
例; 未習未知の知識、異文化発想や概念

上記の広領域誤用分類と平行して、狭領域の語学レベルの誤用タグも以下のように、規範文の基準に必要な語法語彙要素の有無で単純明快に判断できるよう、範疇の見直しを行った。

I 語義的誤用

- a 【母語干渉による代用】日本漢語語彙からの類推

II 語法的誤用

- b 【語順】語順の誤用
- c 【語法成分の不足】必須語法成分の不足
- d 【語法成分の過多】類推などによる語法成分の過多
- e 【定型表現の誤用】定型表現の使用不足と誤用

添削結果は知識データベースに反映され、学習者仮投稿プラットフォームから参照可能となる。このように、最終添削結果

を学習者に返却するだけでなく、学習同士が誤用コーパスから自律的に知識を獲得し、文章作成の早い段階から協調して学習するシステム設計が第二次「写作諸葛」の設計理念である。システム実現のため、学習者の誤用を自動認識し、対応する例文と解説文をリアルタイムで表示する技術も開発中である。

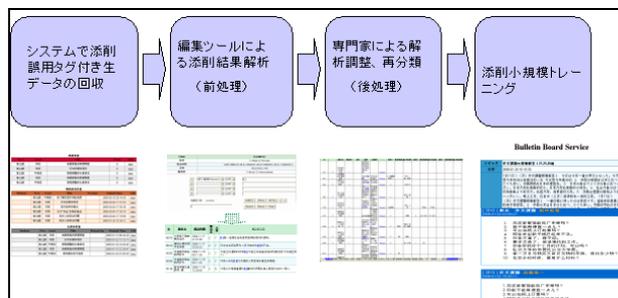


Fig 10: 「写作諸葛」第二次添削システムの流れ

付記

本研究の一部は早稲田大学 2005 年度特定課題研究助成費（一般助成）「日中パラレルコーパスを利用した語学教材開発と授業での活用」（研究代表者：砂岡和子）によるものである。有効評価には李利津、林金錫、林容年(台湾師範大学)、夏維勇(北京大学)、齊慧、嚴金秋(早稲田大学)の協力を得た。引用文献

- [1] 劉松, 砂岡和子, 浦野義頼, 「日中台遠隔中国語口語レッスン管理プラットフォームの開発-早稲田大学 Tutorial Chinese E-Learning Platform」, PC Conference P142-145, 2004
- [2] 劉松, 砂岡和子, 浦野義頼 (2006), 「誤用データ統計機能を備える WEB 中国語作文・添削支援システム」 C&E Vol.020 (Jun. 1, 2006) コンピュータ利用教育協会
- [3] 砂岡和子, 村上公一, 劉松 (2005), 「與評獎和 Self-Editing 相結合進行的網絡華文寫作教學」, International Conference of Internet Chinese Education, Taiwan, P58-66, 2005

参考文献

1. 杉本明子, 「コンピュータ・ネットワークによる作文教育変革の可能性」, 日本教育工学会論文誌, 28(2) : 119-129, 2004
2. 浦崎久美子, 向後千春, 「データベースを利用した作文授業の実践と評価」, 電子情報通信学会技術研究報告, ET, 教育工学, 98(433) : 97-104, 1998
3. 河合和久, 「ネットワーク環境を利用した作文教育」, 情報処理学会研究報告, コンピュータと教育研究会報告, 96(113) : 9-16, 1996