

TNR_Chinese Error Corpus と中国語の誤用研究の インターフェース

関西学院大学 于 康 関西学院大学大学院 田中 良

日本語母語話者の中国語学習者を対象に中国語を教えたり、誤用の研究を行ったりする立場に立つと、文や文章の添削や誤用の傾向性や規則性を見いだすためのデータ処理が必要になってくる。

添削とは、学習者が作った文や文章を直すということであり、レトリックよりも、文法表現の正確さを重視して行われる。データの処理とは、添削した言語データにより詳細な付加情報（タグ）を付与し、タグ付与のコーパスを作成した上で、研究の対象を絞ってキーワードで検索して集計することであり、それによって誤用の傾向性や規則性を見いだすのである。

文や文章の添削については、これまでは、またいまでもそうであるが、朱筆添削がほとんどであるように、手作業に頼らざるを得ない部分が多い。仮に Word の校閲機能を使って添削を行うとしても、添削したものは学習者に返却してしまうだけで、以後のデータの処理にはなかなか結びつかないのが現状であろう。

添削したものは、誤用のデータとして処理する場合、これまでの方法では、手作業によりパソコン上でその誤用のデータを入力しなければならない。タグを付与する場合は、手作業で1つずつ何らかの印を付けたたり書き込みをしたりした上で、目視でその印や書き込まれた結果から誤用の傾向性や規則性を見いだすというのが主な手法であろう。

以上のことが原因であるためか、日本語母語話者の中国語学習者の中国語教育においては、添削が敬遠され、添削済みのデータも生かされないままになっているので、中国語の誤用研究はまだまだ低空飛行を抜けていないといえよう。出現した誤用例を単体で独立させて見た場合、つまり、収集された誤用例を集計し、出現数の多さを加味しない場合、それは必ずしも日本語母語話者の中国語学習者の典型的な誤用とは言えなく、学習者の誤用の傾向性を表すものでもない可能性が高い。場合によっては、非典型例や少数の学習者にしか見られない周辺的な誤用例であるかもしれない。

そこで、この章において、中国語の作文の添削や中国語の誤用研究には何が必要かという現場のニーズを明らかにした上で、それらのニーズに応えるために開発したソフトは何

ができるかを中心に述べていきたい。

1.1 中国語の作文の添削や中国語の誤用研究には何が必要か

作文の添削やデータの処理がいずれも手作業に頼らざるを得ないという現状を打開しなければ、誤用研究はなかなか前には進まず、掘り下げることができない。現状打開を図るために、手作業ではなく、コンピューター上で扱えるソフトを使って中国語の作文の添削や誤用データの処理を行う場合は何が必要なのかについて、学習者や教員、研究者の声を聞かなければならない。つまり、学習者や教員、研究者のニーズに応えるような対策を講じれば、問題解決に結びつくのである。

リサーチした結果、次のような声が聞かれた。

- ①作文を添削するためのソフトがほしい。
- ②添削した内容が自動的に正誤タグの形式に変換されるソフトがほしい。
- ③自動変換された正誤タグの箇所に分析・集計用のタグの付与ができるソフトがほしい。
- ④正誤タグや分析・集計用のタグ付与は、記号ではなく、文字で扱えるソフトがほしい。
- ⑤タグリストからクリックするだけでタグが付与されるソフトがほしい。
- ⑥分析・集計用のタグ付与をする際に、学説別にタグリストを選択できるソフトがほしい
- ⑦タグの定義、典型例、よく使われる語を確認することができるソフトがほしい。
- ⑧タグの判断に迷う場合もあるので、タグを自動判別することができるソフトがほしい。
- ⑨タグの検索や集計ができるソフトがほしい。
- ⑩自動添削や自動タグ付与のソフトがほしい。

⑩については、現在のプログラムの技術ではなかなかクリアできそうもないが、①～⑨なら、少々難しいところがあるとしても、クリアできそうである。

1.2 TNR_ChineseErrorCorpus .WT は何ができるか

①～⑨の課題をクリアするために、私どもは、正誤タグ自動変換機能付きの作文添削ソフト、タグリスト付きの分析・集計用タグ付与ソフト、タグ付与済みのデータを一括して処理できるコンコーダンスの3種類のソフトを開発し、試行錯誤を経て、2013年の秋に以上の3つのソフトを含む TNR_ChineseErrorCorpus を完成した。

TNR_ChineseErrorCorpus は、用途別に2種類ある。

- ①TNR_ChineseErrorCorpus .WT
- ②TNR_ChineseErrorCorpus.WTC

①と②の違いは、前者にはコンコーダンサーTNR_ChineseErrorCorpusConc2.0 が含まれないのに対し、後者には含まれるという点である。コンコーダンサーとは、コーパスからのデータ検索の機能を含む複合的なデータ処理のソフトである。

コンコーダンサーTNR_ChineseErrorCorpusConc2.0 は、多種多様な機能を持つため、本書だけではなかなか説明しきれない。従って、両ソフトの間に見られる因果関係を踏まえ、本書では作文添削ソフトやタグ付与ソフトだけを含む TNR_ChineseErrorCorpus .WT を中心に述べ、コンコーダンサーも含む TNR_ChineseErrorCorpusConc2.0 は、姉妹編として次に刊行することとする。

TNR_ChineseErrorCorpus .WT は、主に次のような機能を持っている。

- ①作文の添削ができる
- ②添削結果はワンクリックで正誤タグの形式に自動的に変換される
- ③自動変換された正誤タグの箇所にワンクリックで分析・集計用の文字タグが付与される
- ④学説別のタグリストからタグを選択し、クリックするだけタグが付与される
- ⑤タグの典型例や定義の確認ができる
- ⑥タグの自動判別ができる
- ⑦検索ができる

以下において、具体例を挙げながら、それぞれの機能を確認してみる。

1.2.1 作文の添削ができる

作文添削ソフト TNR_ChineseWritingCorrection を使って作文の添削ができる。このソフトでは、図 1-1 のように、左側が作文表示区域であり、右側が添削の作業区域である。

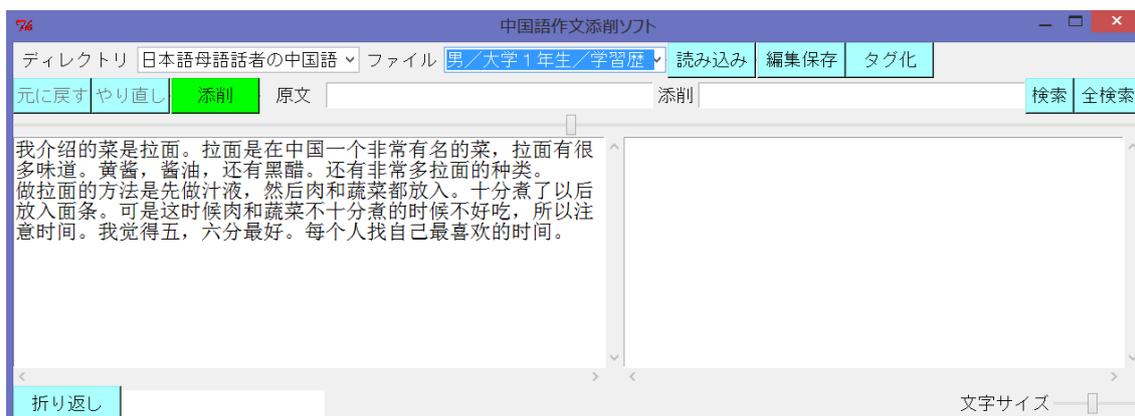


図 1-1

作文の第 1 パラグラフを例にすれば、“我介绍的菜是拉面。拉面是在中国一个非常有名的

菜，拉面有很多味道。黄酱，酱油，还有黑醋。还有非常多拉面的种类。”という文章には、表現のエラー、不使用、過剰使用などといった誤用が容易に観察される。誤用と思われるところを選択して、添削を行えば、図 1-2 になる。右側の作業区域に表記される内容が添削後の結果である。

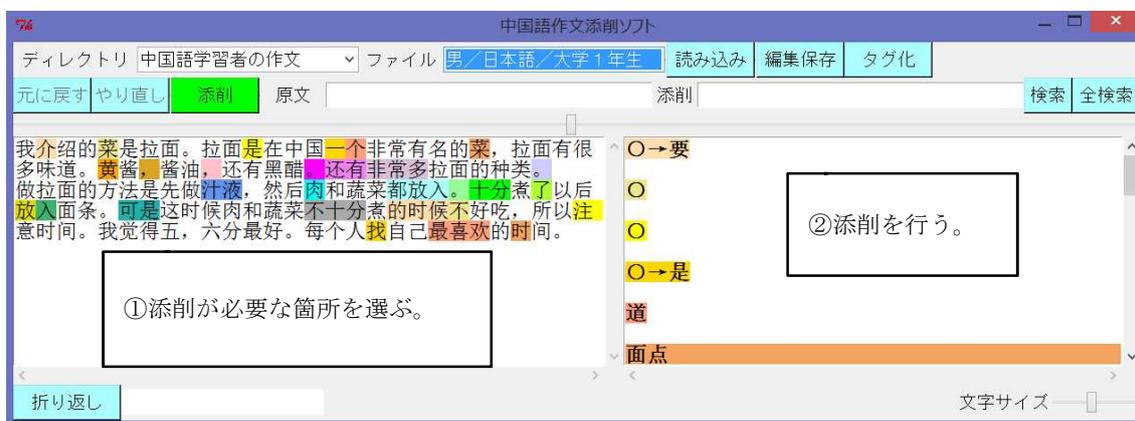


図 1-2

1.2.2 添削結果はワンクリックで正誤タグの形式に自動的に変換される

添削したところは、「誤用→正用」という形で、正誤のタグに自動的に変換し付与される。正誤タグの付与は手作業ではなく、ワンクリックで、図 1-3 のように自動的に行われる。

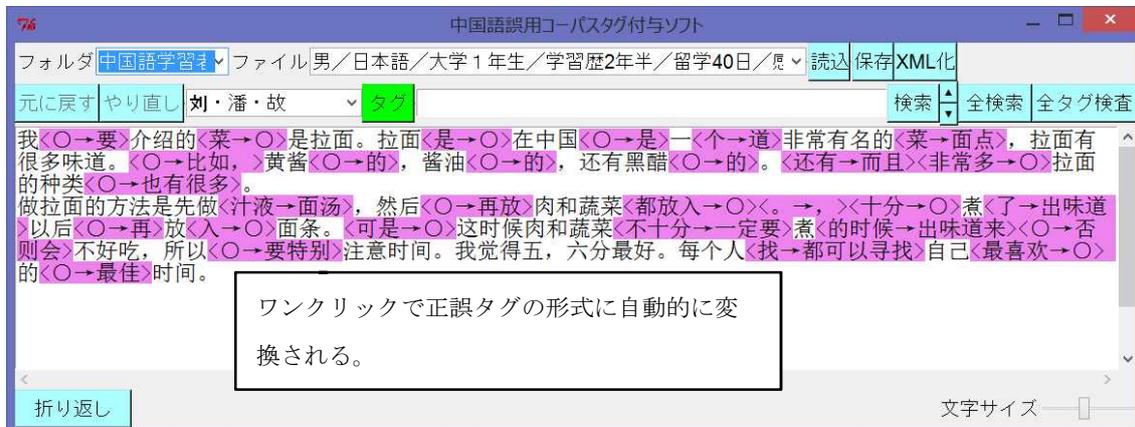


図 1-3

1.2.3 自動変換された正誤タグの箇所にワンクリックで分析・集計用の文字タグが付与される

正誤タグの付与により、どこがどのように間違ったか、どこがどのように直されたかを確認することができるが、それだけでは誤用の傾向性をすぐに見いだすための研究材料にはならない。誤用の傾向性を見いだすためには、分析・集計用のタグを付与しなければならない。分析・集計用のタグとは、統語レベルのタグや意味レベルのタグのことである。

このソフトは、正誤タグの箇所にさらに分析・集計用のタグを付与することができる。

図 1-4 のようにタグリスト付きのタグ付与ソフトを使用すれば、記号ではなく、文字で表示される分析・集計用のタグが付与できる。

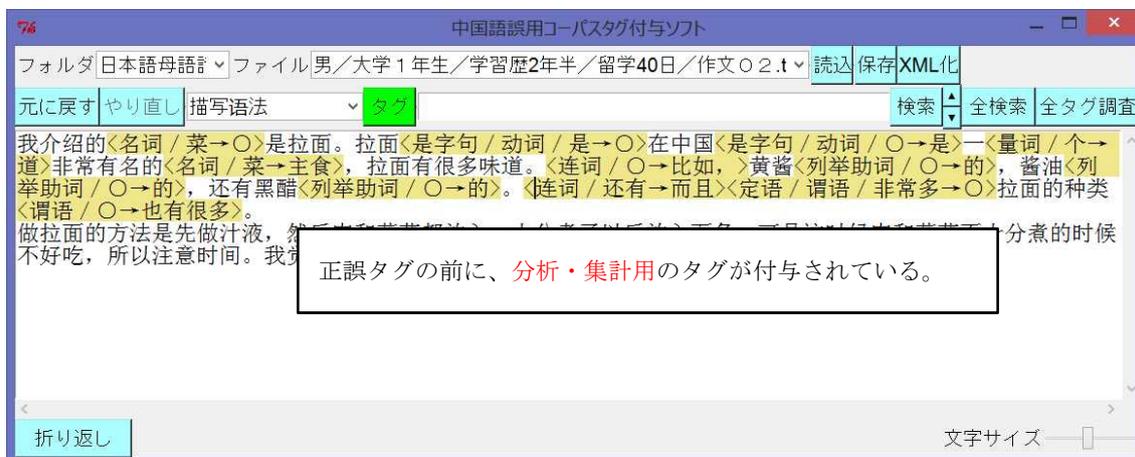


図 1-4

タグ付与後の図 1-4 の内容を書き写すと、次のようになる。

我介绍的<名词/菜→○>是拉面。拉面<是字句/动词/是→○>在中国<是字句/动词/○→是>一<量词/个→道>非常有名的<名词/菜→主食>，拉面有很多味道。<连词/○→比如，>黄酱<列举助词/○→的>，酱油<列举助词/○→的>，还有黑醋<列举助词/○→的>。<连词/还有→而且><定语/谓语/非常多→○>拉面的种类<谓语/○→也有很多>。
做拉面的方法是先做汁液，然后和面并煮熟。面条煮熟后，再放入汤汁中煮熟。分煮的时候不好吃，所以注意时间。我

下線部が、分析・集計用のタグである。“名词”と記したものが名詞使用の間違い、“连词”と記したものが接続詞使用の間違いであり、それぞれの誤用の内容をタグとしてある。

1.2.4 学説別のタグリストからタグを選択し、クリックするだけタグが付与される

タグの分類は学者または学説によって異なっていて、使用者が個別に説を採用しているは一貫性がなくなるので、一貫性を保つためには、ぶれのないタグの設定が求められる。

TNR_ChineseErrorCorpusTagger3.0 のタグリストは、2種類ある。

- ①刘月华・潘文娉・故驩共著の《实用现代汉语语法增订本》（商務印書館 2001）に基づいて作ったタグリスト
- ②张斌主編の《现代汉语描写语法》（商務印書館 2010）に基づいて作ったタグリスト

2種類のタグリストはどちらを使用するかが選択できるので、タグが混ざることなく、タグの一貫性も保たれる。図 1-5 は、《实用现代汉语语法增订本》（「刘・潘・故」と略す）に基づくタグリストであり、図 1-6 は、《现代汉语描写语法》（「描写语法」と略す）に基づくタグリストである。

		タグ付与							自動判別		タグ確認	
学習用												
词类		名词	代词	数词	量词	动词	形容词	副词				
词类		介词	连词	助词	象声词	叹词	复合词					
语法		主语	谓语	宾语	定语	状语	补语	复指	插说			
句法		主谓句	是字句	有字句	连动句	兼语句	存现句	把字句	被字句			
句法		是的句	疑问句	反问句	回声问句	祈使句	比较句	非主谓句				
句法		复句	紧缩句	语序								
研究用												
名词		前缀	后缀	方位词	处所词	时间词						
代词		人称代词	指示代词	疑问代词	特殊代词							
数词		整数	分数小数倍数	概数	序数							
量词		名量词	动量词									
动词		及物动词	不及物动词	零价动词	一价动词	二价动词	三价动词					
动词		动作动词	状态动词	关系动词	能愿动词	持续性动作动词	非持续性动作动词	自主动作动词	非自主动作动词			
动词		体宾动词	谓宾动词	带主谓短语宾语的动词	双宾动词	动词重叠						
形容词		一般形容词	非谓形容词	性质形容词	状态形容词	正向形容词	负向形容词	形容词重叠				
副词		时间副词	范围副词	重复和频率副词	程度副词	语气副词	肯定副词	否定副词	情态副词			
介词		空间介词	时间介词	对象介词	依据介词	缘由介词	其它介词					
助词		结构助词	动态助词	语气助词								
象声词		定型象声词	非定型象声词									
复合词		并列复合词	偏正复合词	动补复合词	动宾复合词	主谓复合词	复杂复合词					
主语		名词主语	代词主语	数词数量词主语	动词主语	形容词主语	的字主语	主谓主语				
谓语		动词谓语	形容词谓语	名词谓语	主谓谓语							
宾语		对象宾语	结果宾语	工具宾语	处所方向宾语	目的原因宾语	存在出现消失宾语	直接宾语	间接宾语			
定语		限定性定语	描写性定语	多项定语								
限定性定语		数量定语	时间定语	处所定语	归属领属定语	范围定语						
描写性定语		性质状态定语	人事物特点定语	用途来源定语	质料定语	职业定语	其它定语					
多项定语		并列关系多项定语	递加关系多项定语	交错关系多项定语								
状语		描写性状语	非描写性状语	多项状语								
多项状语		并列关系多项状语	递加关系多项状语									
补语		结果补语	趋向补语	可能补语	情态补语	程度补语	数量补语	介词短语补语				
复指		重叠复指	称代复指	总分复指								
插说		主观想法态度等插说	推测估计插说	意想不到插说	引起对方注意插说	消息来源插说	举例补充说明插说	总括插说				
主谓句		动词谓语句	形容词谓语句	主谓谓语句	名词谓语句							
兼语句		使令义兼语句	称谓定义兼语句	爱憎好恶义兼语句	说明描写兼语句	有没有兼语句	是兼语句					
存现句		存在句	隐现句									
疑问句		是非问句	特指问句	正反问句	选择问句							
祈使句		可能祈使句	否定祈使句									
比较句		比字句	不比句	没有比较句	不如比较句							
非主谓句		无主句	独词句									
复句		联合复句	偏正复句	多重复句								
联合复句		并列复句	承接复句	递进复句	选择复句							
偏正复句		因果复句	转折复句	条件复句	假设复句	让步复句	取舍复句	目的复句	时间复句			

图 1-5 「刘·潘·故」

タグ編集								
タグ付与								
自動判別								
タグ確認								
学習用								
词类	名词	数词	量词	代词	动词	形容词	区别词	副词
词类	连词	介词	助词	语气词	象声词	感叹词	助动词	
语法	短语	主语	谓语	宾语	中心语	定语	状语	补语
句法	名词性谓语句	形容词性谓语句	动词谓语句	述补谓语句	述宾谓语句	连动谓语句	兼语谓语句	主谓谓语句
句法	名词性非主谓句	动词性非主谓句	形容词性非主谓句	象声词叹词非主谓句	疑问句	祈使句	感叹句	双宾语句
句法	把字句	被字句	得字句	有字句	存现句	比字句	是字句	连字句
句法	可逆句	联合复句	偏正复句	多重复句	紧缩复句			
研究用								
名词	专有名词	个体名词	集体名词	实体名词	物质名词	具体名词	抽象名词	无价名词
名词	有价名词	有量名词	非量名词	关系名词	有序名词	时间名词	处所名词	方位词
数词	系数词	位数词	概数词					
量词	物量词	动量词	时量词	复合物词				
代词	人称代词	指示代词	疑问代词					
动词	动作动词	心理动词	使令动词	存现动词	形式动词	关系动词	判断动词	趋向动词
动词	不及物动词	黏宾动词	双宾动词	一般动词	零价动词	一价动词	二价动词	三价动词
形容词	动态形容词	静态形容词	量度形容词	非量度形容词	状态形容词			
区别词	单纯区别词	派生区别词	复合区别词					
副词	时间副词	程度副词	范围副词	语气副词	否定副词	情态副词		
连词	联合连词	偏正连词						
助词	结构助词	时态助词	比况助词	表数助词	列举助词	限定助词		
短语	偏正短语	述补短语	述宾短语	主谓短语	联合短语	同位短语	连动短语	兼语短语
短语	量词短语	方位短语	介词短语	的字短语				
主谓句	名词性谓语句	动词性谓语句	形容词性谓语句	主谓谓语句				
动词性谓语句	动词谓语句	述宾谓语句	述补谓语句	连动谓语句	兼语谓语句			
非主谓句	名词性非主谓句	动词性非主谓句	形容词性非主谓句	象声词叹词非主谓句				
疑问句	是非问句	选择问句	正反问句	特指问句	反问句			
是非问句	语调是非问句	吗问句	吧问句	特指性是非问句				
祈使句	命令句	禁止句	建议句	劝阻句	请求句	乞免句		
存现句	表存在的存现句	表出现的存现句	表消失的存现句					
是字句	表判断是字句	表强调是字句						
可逆句	供用义可逆句	存现义可逆句	合并或烹调义可逆句	是等于像字句等可逆句	对朝挨靠等加着的可逆句	表纯粹计量关系的可逆句	气象类可逆句	
双宾语句	交类双宾语句	教类双宾语句	写类双宾语句	接类双宾语句	借类双宾语句	问类双宾语句	欠类双宾语句	称类双宾语句
双宾语句	放类双宾语句	高类双宾语句						
联合复句	并列复句	连贯复句	递进复句	选择复句	取舍复句	解释复句		
偏正复句	因果复句	目的复句	条件复句	假设复句	转折复句	让步复句		
宾语的类型	结果宾语	工具宾语	方式宾语	受事宾语	处所宾语	时间宾语	原因宾语	
宾语的类型	目的宾语	对象宾语	角色宾语	施事宾语	使动宾语			
论元的分类								
主事	施事	致事	经事	系事	起事			
客事	受事	成事	使事	感事	涉事	止事	位事	任事
与事	当事	共事						
补事	补事							
可有成分	凭事	境事	因事	关事	比事			

図 1-6 「描写语法」

タグリスト内のタグは、さらに、

- ① 「学習用」のタグ
- ② 「研究用」のタグ

に2分類している。「学習用」のタグは、下位分類を含まない上位の分類のみで、中国語学を研究対象としない学習者や中国語教員のためのタグである。例えば、「描写语法」の品詞分類では、“名词”や“动词”のように、上位の分類にとどまり、名詞の下位分類や動詞の下位分類は設けていない。

それに対し、「研究用」のタグは、更に下位分類を詳しく記した分類で、中国語学を研究対象とする学習者や研究者のためのタグである。例えば、“名词”という上位分類ではなく、“专有名词”“个体名词”“集体名词”“实体名词”“物质名词”“具体名词”“抽象名词”“无价名词”“有价名词”“有量名词”“非量名词”“关系名词”“有序名词”“时间名词”“处所名词”“方位词”のように、より詳細な下位分類のタグを使う。これによって、“名词”の誤用において、どのような種類の名詞がよく間違われたかを詳細に観察することができるので、他の種類の名詞の誤用と比べながら、名詞誤用の規則を見いだすことも可能になる。

1.2.5 タグの典型例や定義を確認することができる

特に、「研究用」のタグにおいては、下位分類が細かく分類されているので、タグの種類も多く、どのタグをどの場合に使うべきか判断するのが難しい。そのため当該する語の典型例を確認することができれば、非常に助けになる。そこでこのソフトでは、確認したいタグをクリックすれば、そのタグの典型例が表示されるという機能を設けている。

例えば、図 1-7 のように、“语气副词”をクリックすれば、「语气副词」リストによって“语气副词”の典型例が表示され、確認することができる。

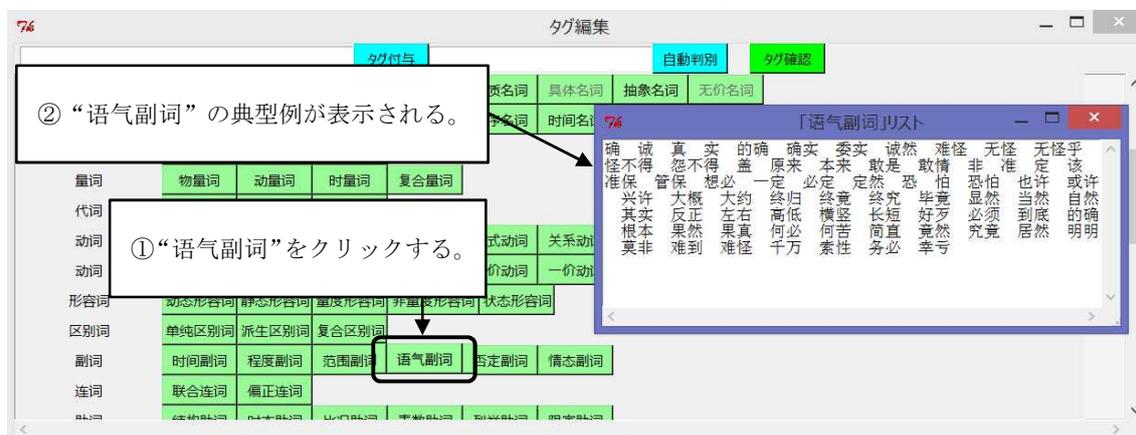


図 1-7

タグの定義を確認したい場合もある。図 1-8 のように、予め定義が記録されているタグをクリックすれば、タグの定義、用例、典型例が表示される。

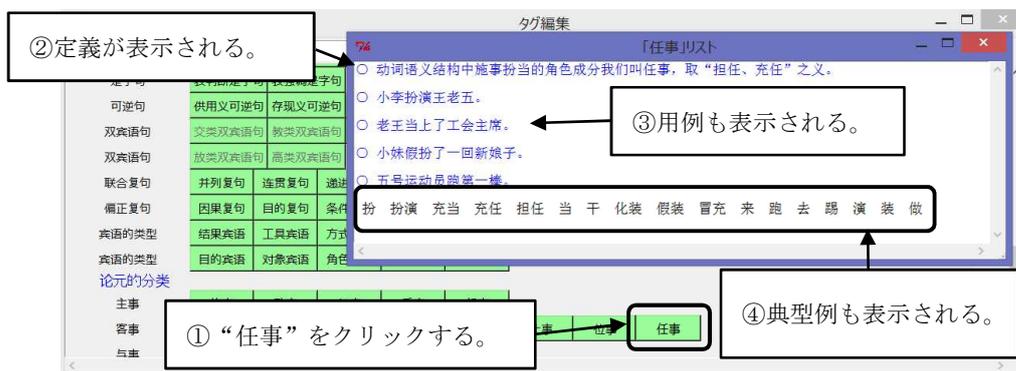


図 1-8

タグの定義、用例、典型例の表示は、すべてのタグが備えているわけではない。つまり、確認を行っても、定義だけ、用例だけ、典型例だけや、または、定義と用例だけ、定義と典型例だけなどが表示される場合もあるということである。

1.2.6 タグを自動判別することができる

付与したいタグは、タグリストの中から選択し、クリックするだけで付与することができるというのがこのソフトの特徴の 1 つである。しかしタグの選択は、使用者が判断するため、思い違いで選択ミスが生じてしまう可能性も否定できない。自動的にタグを判別する機能があれば、さらに手間を省くことができるだけでなく、人為的な選択ミスの軽減にもつながるであろう。タグ自動判別機能とは、タグを付与しようとする語を選択、またはその語を入力するだけで、付与されるべきタグが自動的に表示される機能のことである。

例えば、“过”にタグを付与したい。しかし、“过”はどのように分類され、どのようなタグを付与すればよいかを迷ってしまう場合、普通なら文法書を調べなければならない。それでは非常に手間暇がかかるため、タグ付与の作業はなかなかかどらない。

このソフトは、図 1-9 のように、“过”と入力して、「自動判別」のボタンをクリックすれば、“趋向动词”“二价动词”“程度副词”“时态助词”“高类双宾语句”“位事”の 6 つのタグの選択肢が自動的に表示され、この中から選択できるようになっている。これで作業の手間が省けるし、タグの判断ミスのゼロ化にもつながる。



図 1-9

1.2.7 検索ができる

タグリストにない語はタグ自動判別機能が使えないため、現れるたびにタグを判断して付与することになる。その際、これまで作業中にどのようなタグが付与されてきたかを確認することができれば、タグ付与の一貫性が保たれる。また、高度な検索・集計機能を備えていなくても、正誤タグに変換された誤用例から特定の誤用を探し出し誤用の傾向性を即座に確認することができれば、すぐに授業に還元することができるであろう。

例えば、“再”について、これまでどのようなタグが付与されたかを確認することとする。

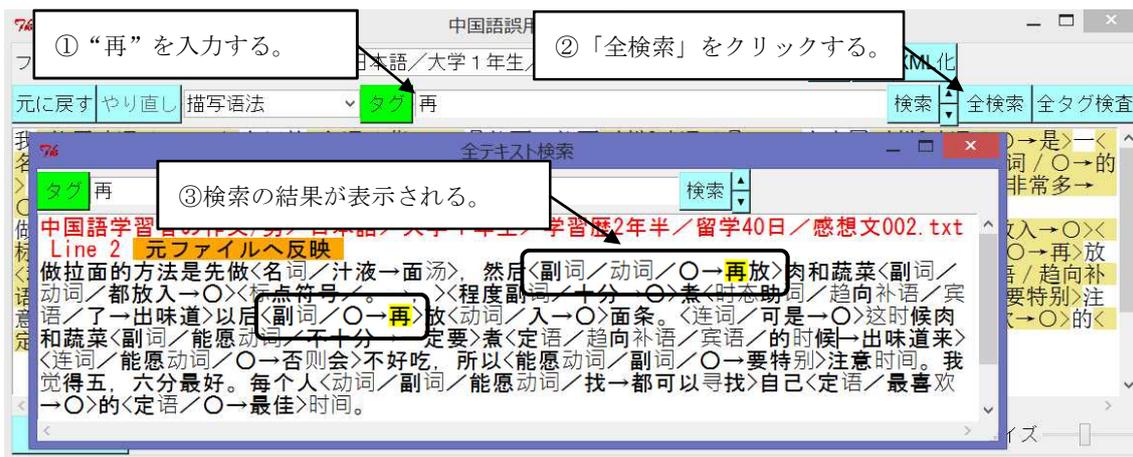


図 1-10

図 1-10 のように、“再”と入力してから、「全検索」ボタンをクリックする。「全検索」とは、フォルダーに保存されたファイルすべてを対象に検索を行うことである。すると検索の結果が表示される。例として取り上げている作文はワンパラグラフだけなので、データの量が少ないが、“再”に付与されたタグは2箇所あり、いずれも黄色の字で表示される。確認した結果、“再”には、いずれも“副词”というタグが付与されている。つまり、“再”なら、“副词”というタグを付与すればよいということがわかったのである。

次は、過剰使用の誤用を調べる。過剰使用のタグは「→○」という形で示されているので、この「→○」をキーワードとして検索すれば、過剰使用の実態は確認できる。

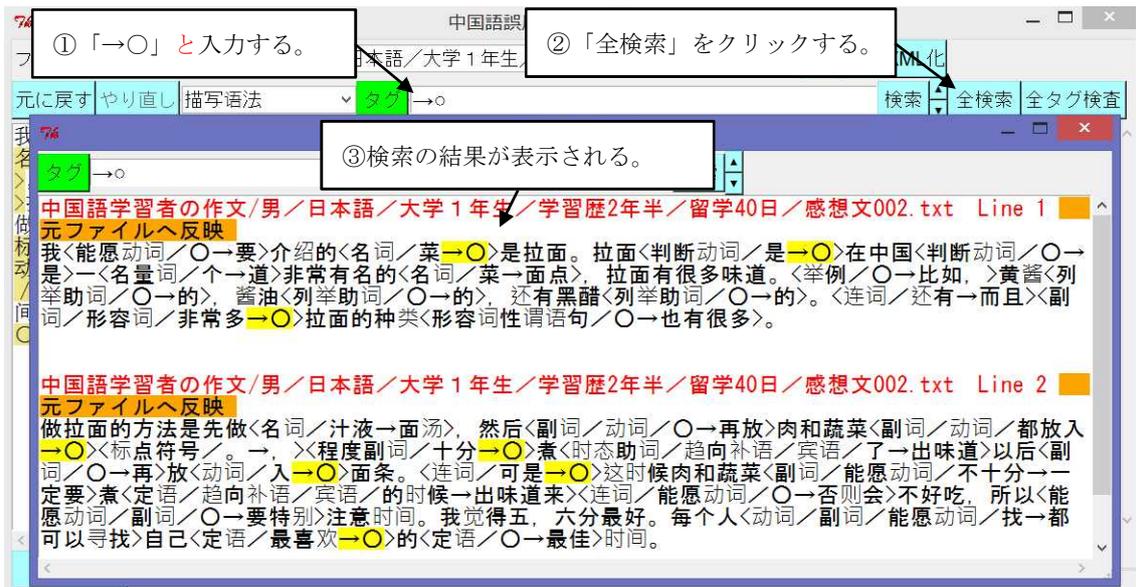


図 1-11

図 1-11 のように、「→○」と入力してから、「全検索」ボタンをクリックする。「→○」の部分はいずれも黄色になって表示される。ここの検索機能は簡易機能なので、集計などはできないが、目視により、どのような場合に「→○」という誤用がよく見られるかを確認することができる。もしこの「どのような場合」を集計しその中から規則性を見いだそうとするなら、コンコーダンスーを使わなければならない。これについては、第 7 節で詳しく述べる。

1.3 まとめ

この章において、中国語の作文の添削や中国語の誤用研究には何が必要か、や TNR_ChineseErrorCorpus.WT は何ができるのかを中心に述べてきた。

TNR_ChineseErrorCorpus.WT にある機能をまとめてみると、次のようになる。

- (1) 作文の添削
- (2) 添削の内容をワンクリックで正誤タグの形式に自動的に変換
- (3) 自動変換された正誤タグの箇所にワンクリックで分析・集計用の文字タグを付与
- (4) 学説別のタグリストからタグを選択し、クリックするだけタグを付与
- (5) タグの典型例や定義の確認
- (6) タグの自動判別
- (7) 検索

以上の機能を備えているソフトであるが、これまで述べてきたように、現場からのニーズを踏まえた上で制作したものなので、使い勝手については今後更なる改善をしなければならないところがあるが、手作業で行うよりは、遥かに便利なツールになっていると言え

よう。

特に、TNR_ChineseErrorCorpus.WT 中の TNR_ChineseWritingCorrection は、単独で作文の授業にも使うことができるので、使いこなせば、中国語学の研究だけでなく、中国語の教育にも大いに役立つものになるであろう。