

上級中国語学習者誤用コーパス (満月コーパス) の構築

張學博 (東京外国語大学大学院生)

上級中国語学習者誤用コーパス (満月コーパス)

▶ コーパスの特徴

- ・ オンライン上での一般公開を目指す。
- ・ 分類された誤用タイプ、学習者の背景、作文テーマ等から検索可能である。
- ・ 添削作文並びにもと作文の両者が閲覧可能である。

▶ サンプル

18	...父母打电话, 回老家的时间, 我觉	想法 留心	也是孝敬父母的一个好方法。第三...
学習者ID	該当部分前方	該当部分	該当部分後方

1 to 10 (21)

カテゴリから検索

☐ 复句的固定句式【一...就】 ☐ 复句 ☐ 名词 ☐ 副词 ☐ 动词 ☐ 主语 ☐ 对象介词 ☐ 原因介词 ☐ 形容词 ☐ 指示代词 ☐ 因果复句: 因为...所以 ☐ 动态助词 ☐ 结构助词 ☐ 构造助词 ☐ 并列复句: 既...也 ☐ 否定副词 ☐ 不及物动词 ☐ 时间名词 ☐ 固定句式: 无论...也 (都) ... ☐ 能愿动词 ☐ 趋向补语 ☐ 能愿动词 ☐ 趋向补语 ☐ 连词 ☐ 【表现】 动词: 和...商量 ☐ 表现 ☐ 因果复句: 既然...就 ☐ 介词短语: 在...里 ☐ 【表现】 动词: 从...开始

学習者情報から検索

18	...父母打电话, 回老家的时间, 我觉	想法 留心	也是孝敬父母的一个好方法。第三...
学習者ID	該当部分前方	該当部分	該当部分後方

1 to 10 (21)

カテゴリから検索

学習者情報から検索

TOEFL (IBT): 0点 - 120点

学習期間: 0年 - 20年

海外滞在経験: ☐ あり ☐ なし

性別: ☐ 男 ☐ 女

年齢: ☐ 10代 ☐ 20代 ☐ それ以上

国籍: ☐ 日本 ☐ 中国 ☐ その他

母語: ☐ 日本語 ☐ 中国語 ☐ 英語 ☐ その他

教育言語: ☐ 日本語 ☐ 中国語 ☐ 英語 ☐ その他

中国語学習者中国語作文誤用コーパス検索サイト

サイトについて

本サイトは、中国語学習者コーパスのデータを検索できるWebアプリケーションです。コーパスデータは、東京外国語大学・中国語専攻・望月圭子研究室、佐野洋研究室、中聖歌、福地潤岡 氏らの協力のもと収集されました。検索可能なデータ件数は、平成25年10月〇日時点で〇〇件です。

誤用文字列: ⇒ 修正文字列: 検索 リセット

タイプ検索: ☐ 置換 ☐ 削除 ☐ 追加

1 to 10 (21)

学習者ID	該当部分前方	該当部分	該当部分後方
18	...东京。东京的生活费用比我老家太	贵	为减轻那个负担, 我现在打工挣钱...
18	...我老家贵, 所以父母的负担很大,	为了	减轻那个负担, 我现在打工挣钱。...
18	...贵, 所以父母的负担很大, 为减轻	那个负担	负担, 我现在打工挣钱, 我挣的钱...
18	...自己打工挣一些钱, 我们的负担	减少	了。"第二个是时间的方面, 上大...
18	...每天很忙, 没有时间, 所以我没有	对	父母打电话, 或者去老家的时间。...
18	...啊, 所以我没有对父母打电话,	或者他	去老家的时间, 但是, 毕业大学。...
18	...啊, 所以我没有对父母打电话或者	去	老家的时间, 但是, 毕业大学, 弄...
18	...以, 我现在改变了自己想法, 无论	如何	忙, 也通过对父母打电话回老家的时...
18	...变了自己想法, 无论如何忙, 也排	除	父母打电话, 回老家的时间, 我觉...
18	...父母打电话, 回老家的时间, 我觉	想法 留心	也是孝敬父母的一个好方法。第三...
学習者ID	該当部分前方	該当部分	該当部分後方

作業の流れ

▶ コーパスデータ

- ・ 承諾書, 学習者情報
- ・ 作文の収集
- ・ タグリスト

▶ コーパス構築作業

- ・ 添削
- ・ タグ付け
- ・ XML化

コーパスデータ：承諾書・学習者情報

- ▶ 最終的に作文を公開するにあたり、学習者からは必ず「承諾書」へのサインを得ている。

▶ 学習者情報

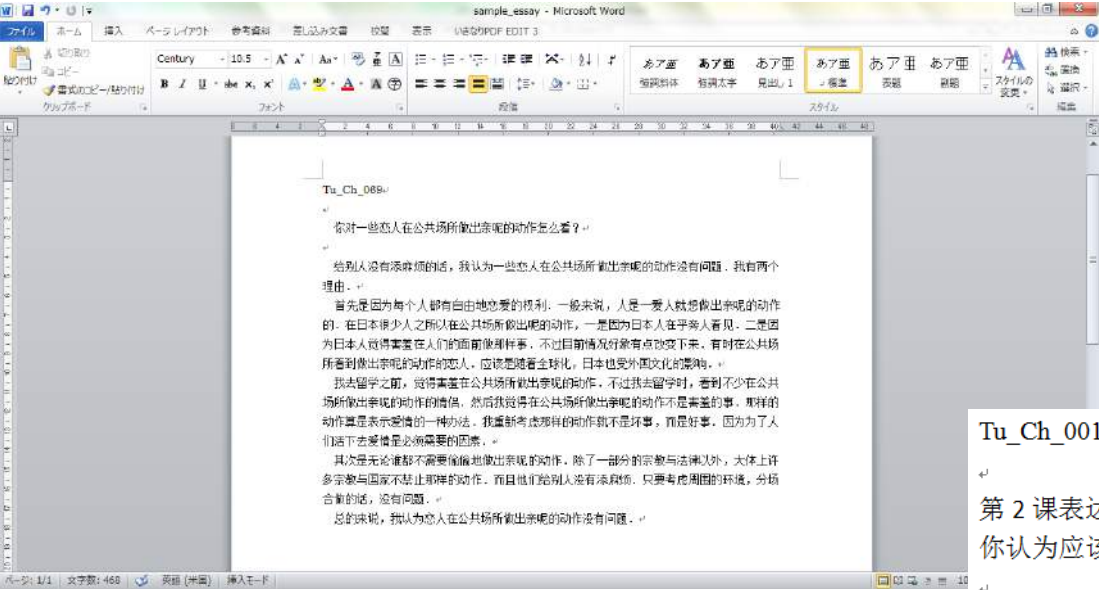
- ・ 国籍 ・ 母語 ・ 教育言語 ・ 中国語学習歴 ・ 中国語学習機関
- ・ 留学経験（期間，機関，所在地） ・ 家庭で話す言葉（父，母，兄弟）
- ・ 友人と話す言語 ・ 小学校の教育言語 ・ 中学校の教育言語
- ・ 高校の教育言語 ・ 英検 ・ TOEFL(iBT) ・ TOEIC ・ IELTS(academic)
- ・ IELTS(general) ・ 中国語検定 ・ HSK(筆記) ・ HSK(口頭)

コーパスデータ：作文の収集 (作文の概要)

- ▶ 東京外国語大学中国語学科の 2 年生から 4 年生までの作文を収集。
 - ▶ 学習者は、全体で 81 名（2 年生：38 名，3 年生：26 名，4 年生17 名）。
 - ▶ 現在収集した作文の数は、306 本（2013 年度前期のみ）。
 - ▶ 文字数は、約 102,000 字。
-
- ▶ 現在、Moodle システム (a type of software e-learning platform, a Learning Management System / Virtual Learning Environment) を通して、学習者から作文を収集し、ワードのコメント機能で添削を行い、学習者に返却する。
 - ▶ 直接的にコメント機能で行った添削は、コーパス構築には反映されないが、学習者にとって非常に有益である。

作文のサンプル

もと作文



添削作文

Tu_Ch_001

第 2 课表达练习 3-3
你认为应该怎样孝敬父母？

我认为孝敬父母时，**我** 要表示三个方面的感谢。第一个方面 **是** **关于** 金钱方面的感谢。我老家在群馬，不过现在我住在东京。东京的生活费用比我老家 **太 贵**，所以父母的负担很大。为 **了** 减轻 **那个** **他们的** 负担，我现在打工挣钱。我挣的钱还没 **不** 那么多，不过我妈妈常常说，“**因为** 你现在自己打工挣**了** 一些钱，**所以** 我们的负担减 **少** **少** 了。”第二个是 **关于** 时间 **的** 方面 **的感谢**。上大学以后，我的生活很充实，每天很忙，没有时间。所以我 **既** 没有 **对** 给 父母打电话，**或者 也 没有** **回 去** 老家的时间。”

コメント [潘55]: 【删除】【主语】

コメント [S56]: 动词

コメント [S57]: 介词（对象）

コメント [m58]: 【替换】高很多【程度补语】

コメント [S59]: 介词（原因）

コメント [潘60]: 【替换】他们【指示代词】

コメント [潘61]: 【添加】的【结构助词】

コメント [S62]: 代词

コメント [潘63]: 【删除】【否定副词】

コーパスデータ：タグリスト

▶ 添削用

「削除」(removal)、「添加」(addition)、「替換」(replacement)、
「移動」(transfer) (削除＋添加)

▶ タグリスト

张斌 (2007) 《新编现代汉语》 复旦大学出版社.

齐沪扬 (2012) 《现代汉语》 商务印书社.

齐沪扬 (2010) 《对外汉语教学语法》 复旦大学出版社.

→主に、この三冊の本にあるカテゴリーからタグリストを作成した。

タグは、作業過程で必要に応じて、適宜追加している。

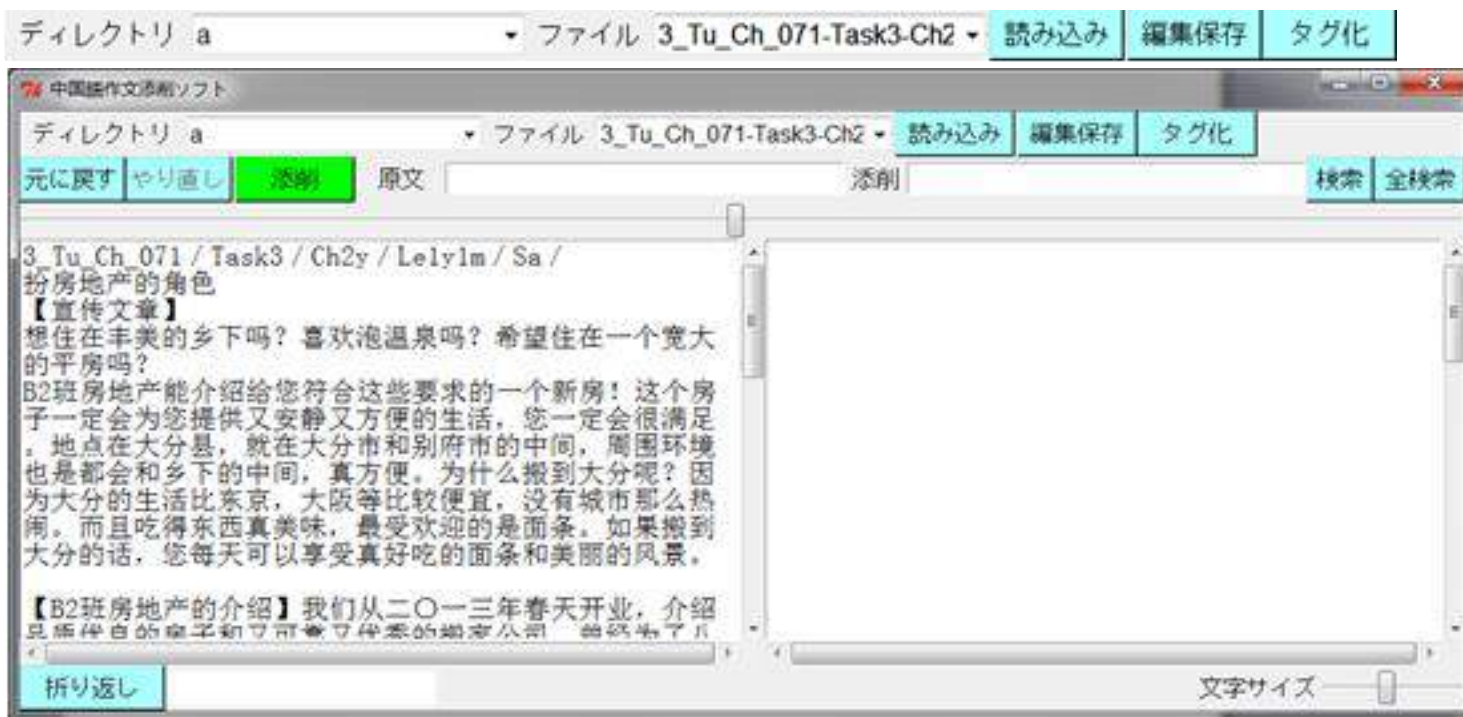
タグリスト：サンプル

74 タグ編集															
タグ付与				自動判別		タグ確認									
名词	专有名词	可量名词	不可量名词	时间名词	处所名词	方位词									
可量名词	个体名词	集体名词	物质名词	抽象名词											
数词	系数词	位数词	系数数词	数量数词	概数词										
动词	动作动词	存现动词	关系动词	能愿动词	趋向动词	使令动词									
动词	及物动词	不及物动词													
动词	静态动词	持续动词	瞬间动词	终结动词											
动词	重叠动词														
形容词	性质形容词	状态形容词													
区别词	区别词														
副词	程度副词	范围副词	时间副词	情态副词	否定副词	语气副词	关联副词								
代词	人称代词	指示代词	疑问代词												
连词	连词														
介词	时间介词	处所介词	方向介词	范围介词	工具介词	依据介词	对待介词	原因介词	目的介词	与事介词	处置介词	被动介词	介绍介词	伴随介词	比较介词
助词	结构助词	时态助词	时制助词	比况助词	表数助词	列举助词	其他助词								
语气词	语气词														
拟音词	叹词	象声词													
短语<短语的结构分类>	主谓短语	述宾短语	述补短语	偏正短语	联合短语	同位短语	联动短语	兼语短语	量词短语	方位短语	介词短语	“的”字短语	“所”字短语	比况短语	
短语<短语的功能分类>	体词性短语	谓词性短语	加词性短语												
主语<语义分类>	施事主语	受事主语	中性主语												
主语<功能分类>	体词性主语	谓词性主语													
谓语	谓语														
宾语	宾语	双宾语													
补语	结果补语	趋向补语	程度补语	情态补语	数量补语										
定语	定语														
状语	状语														
主谓句	名词性谓语句	动词性谓语句	形容词性谓语句												
非主谓句	非主谓句														
陈述句	陈述句														
祈使句	祈使句														
感叹句	感叹句														

コーパス構築作業：添削

中国語誤用コーパスタグ付与ソフト（東京外大望月圭子科研専用）

▶ TNR_ChineseWritingCorrection.exe



▶ 本ソフトウェアで添削できるのは、「替換」「削除」「添加」「移動」の四種類である。

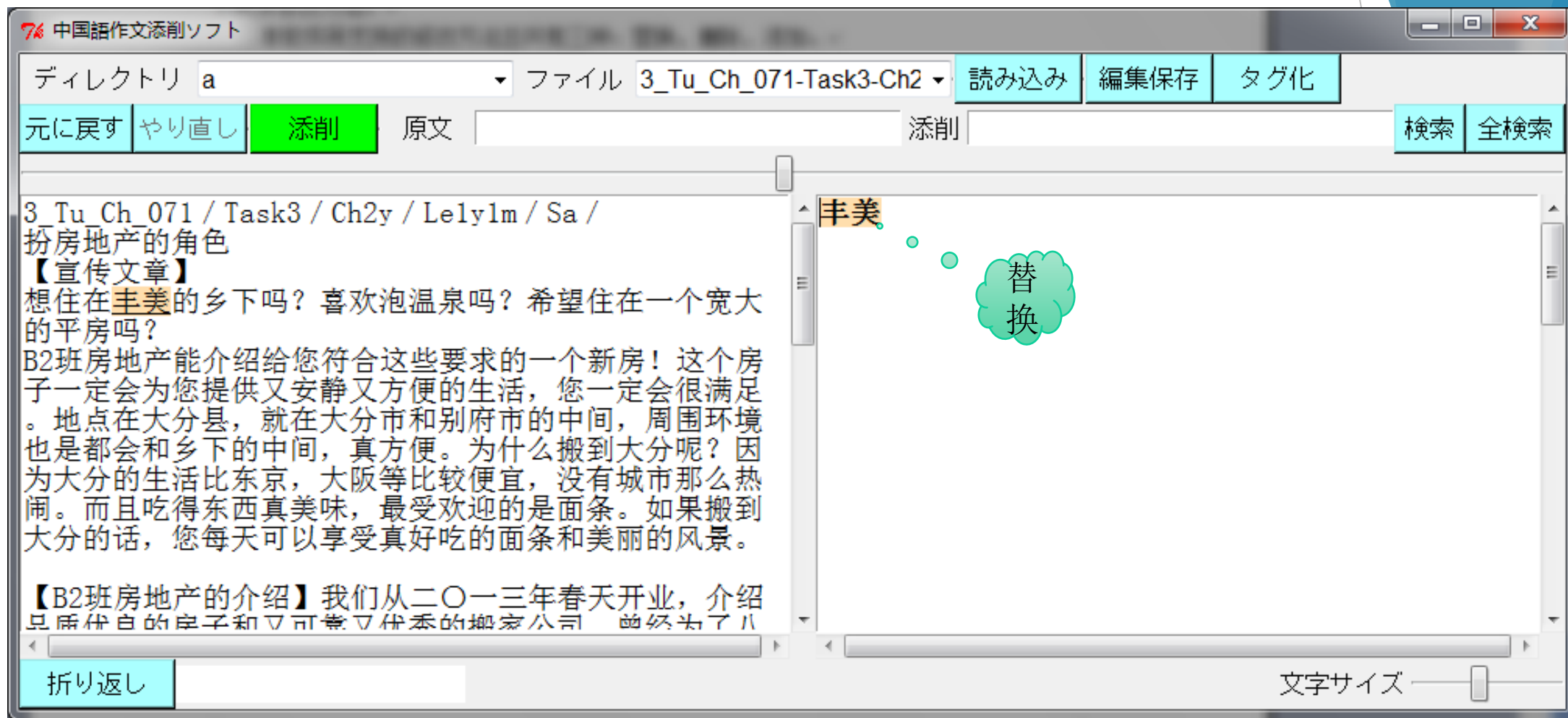
替换_1

3_Tu_Ch_071 / Task3 / Ch2y / Le1y1m / Sa /
扮房地产的角色
【宣传文章】
想住在丰美的乡下吗？喜欢泡温泉吗？希望住的平房吗
B2班房地 给您符合这些要求的一个新子一定会为您提供又安静又方便的生活，您一

誤用箇所をドラッグして、青く表示し、「右クリック」して、【コピー、添削】の表示から、「添削」を選択する。

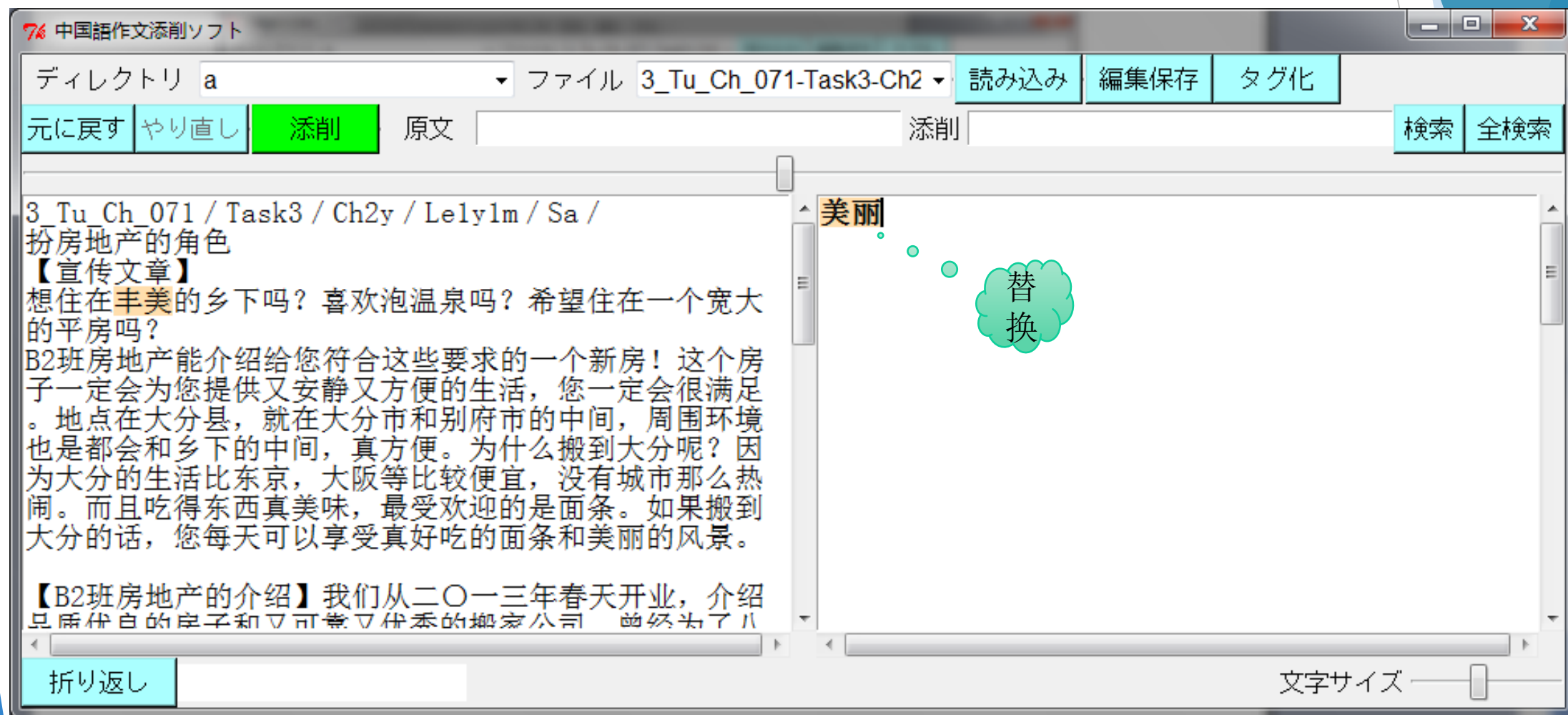
替换_2

右半分の画面に、自動的に選択した語句（〈丰美〉）が表示される。



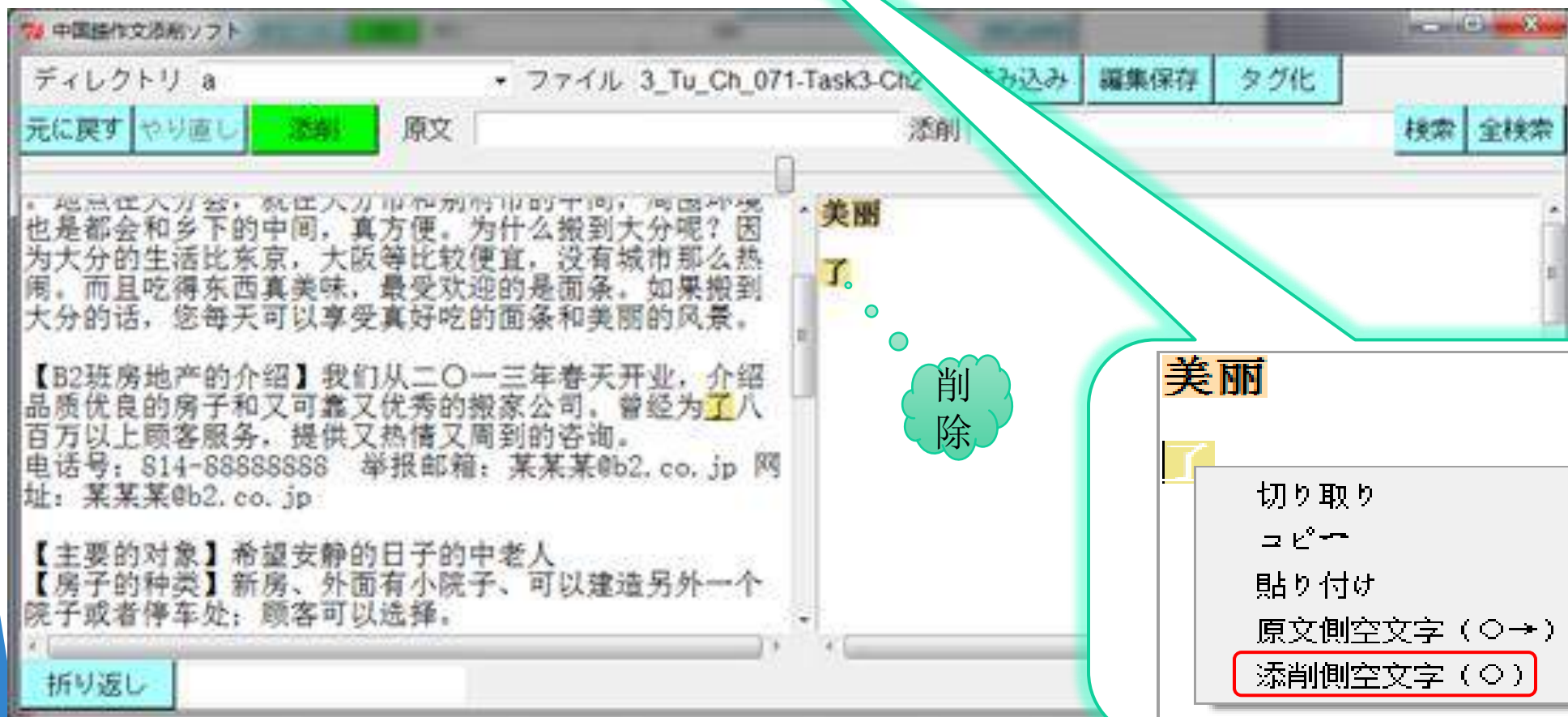
替换_3

右画面に出た〈丰美〉を削除し、直接ふさわしい語句（〈美丽〉）を入力する。



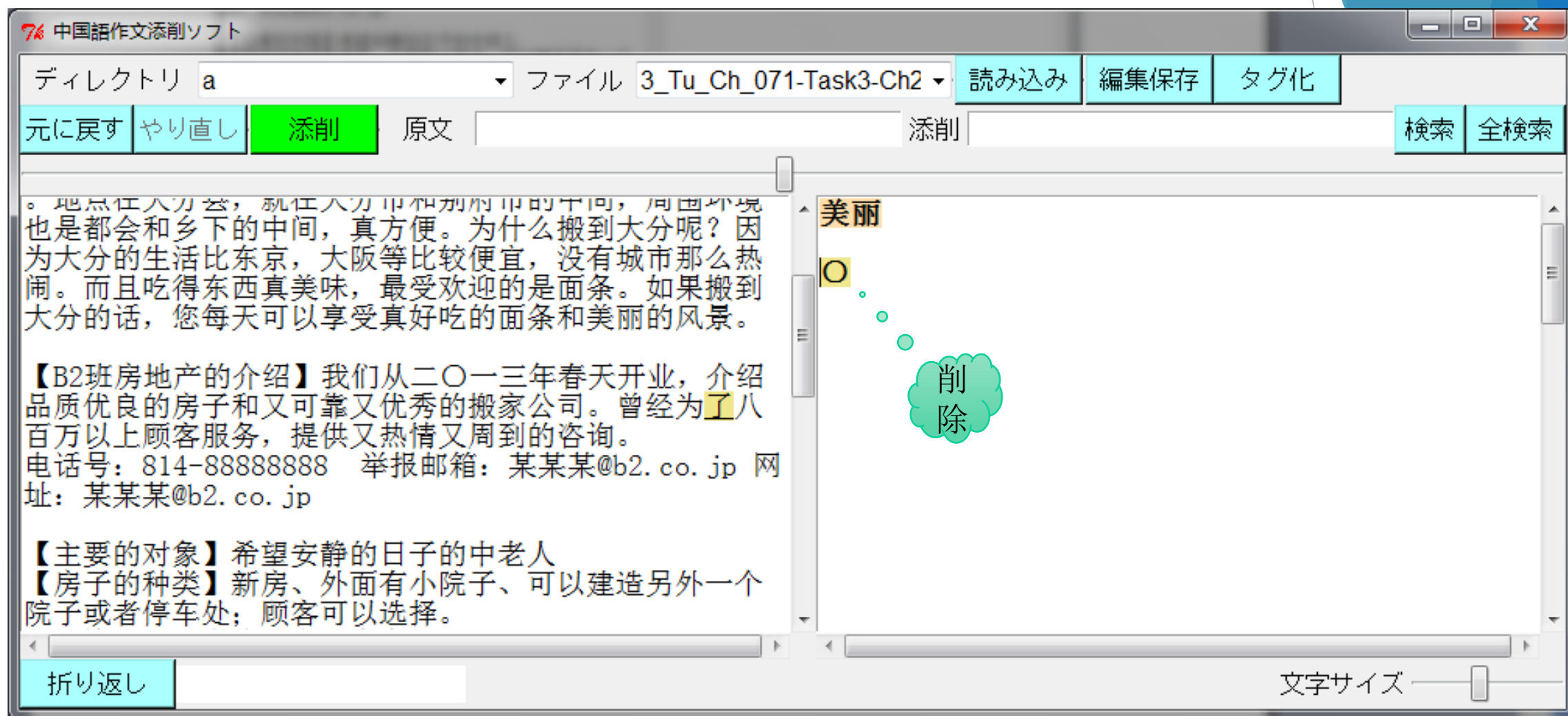
削除_1

誤用箇所をドラッグして、青く表示し、「右クリック」して、5つの選択肢から、「添削側空文字」を選択する。



削除_2

右画面に自動的に「○」が表示され、添削は完了となる。



添加_1

誤用箇所の直後の文字をドラッグして、青く表示し、「右クリック」して、5つの選択肢から、「原文側空文字（○→）」を選択する。



添加_2

右画面に自動的に「○→」が表示され、「→」の直後に付け加えるべき語（〈过〉）を入力する。

美丽

○

○→

添
加



美丽

○

○→过

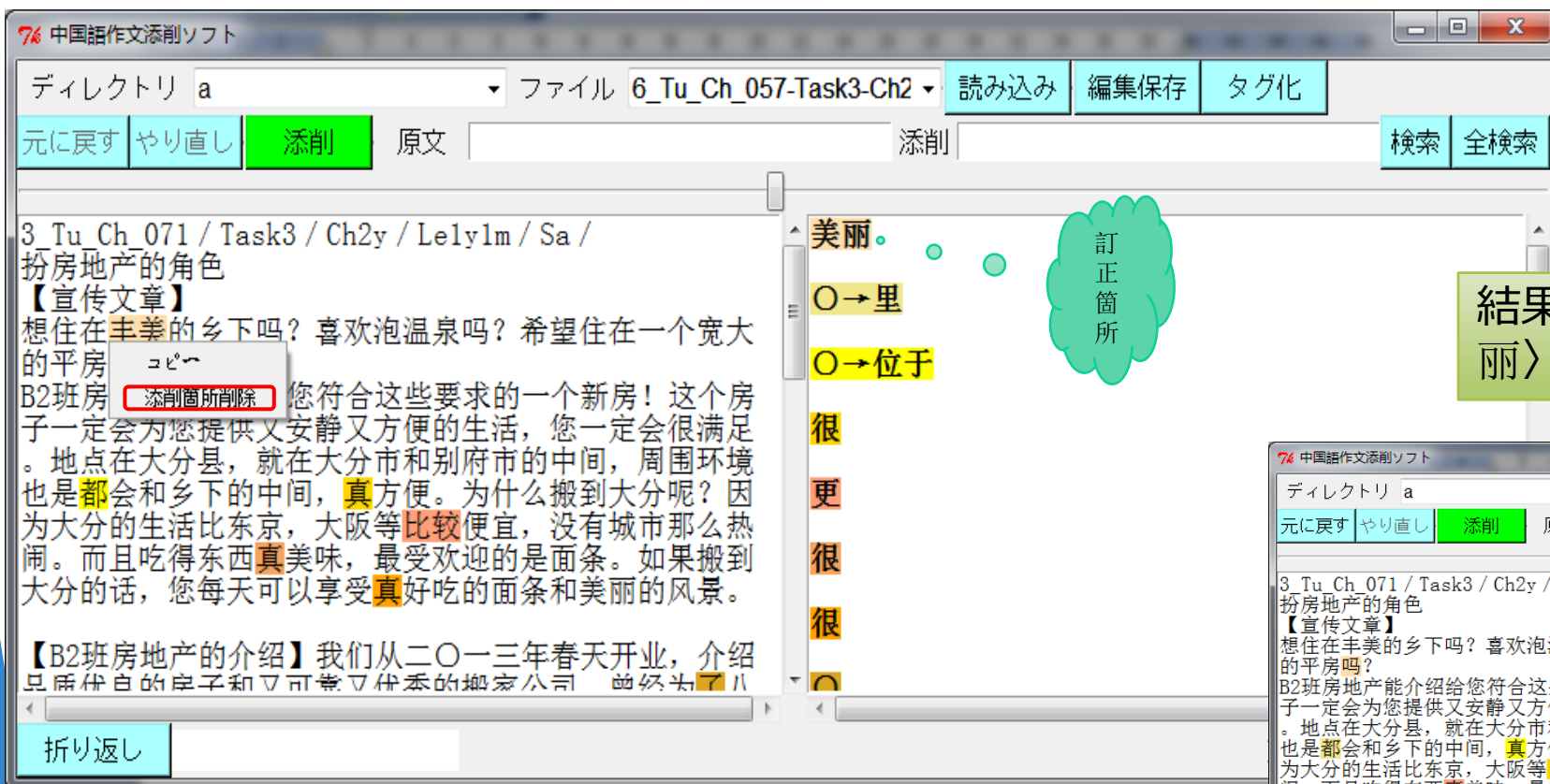
添
加

移動（削除＋添加）

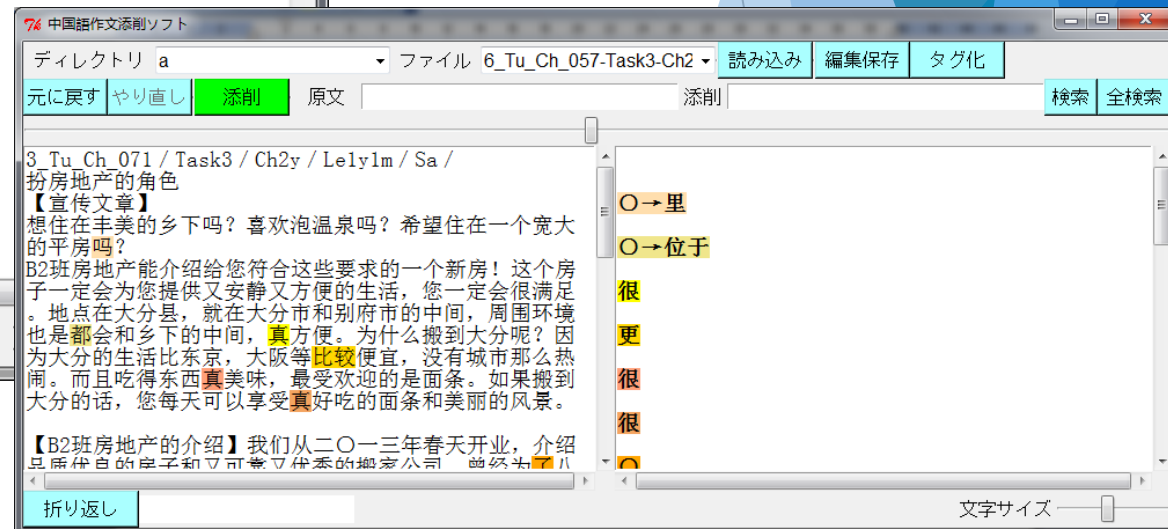
- ▶ 「移動」操作は、直接行えないため、「削除」→「添加」の順で行う。
- ▶ 主に語順の問題であるため、誤用箇所をいったん「削除」し、適切な場所に「添加」する。

添削ミスの解除方法

左画面の解除したい語（〈丰美〉）の直前にカーソルを合わせ、右クリックで「添削箇所削除」を選択する。



結果、右画面の削除したい語（〈美丽〉）が消える。



添削済み作文の保存

読み込み

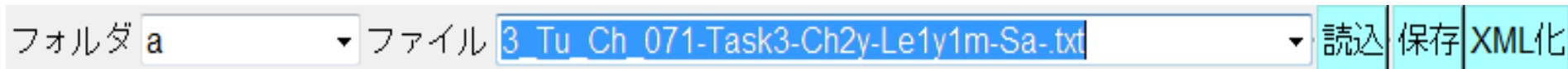
編集保存

タグ化

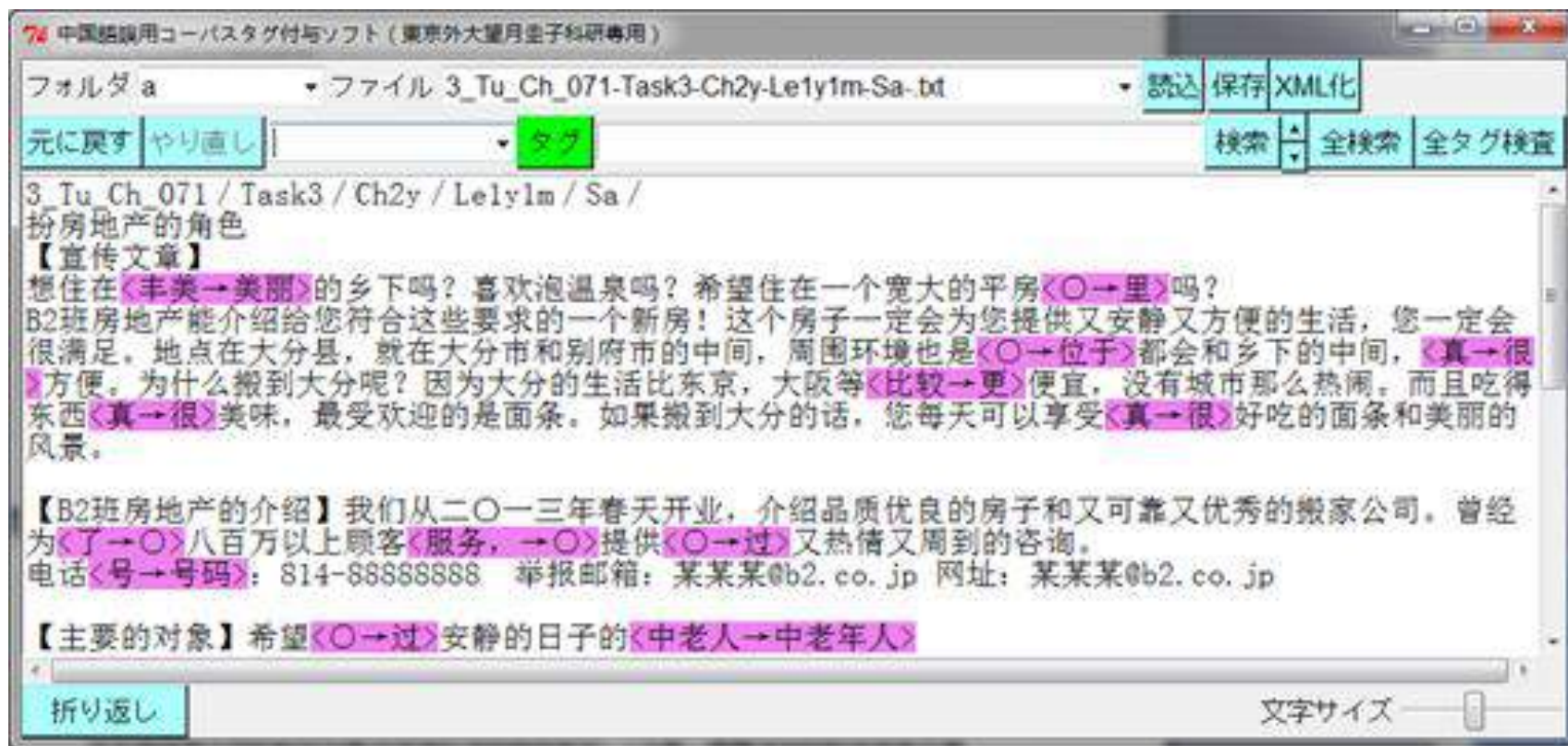
- ▶ Word と同様、作業途中に適宜「編集保存」を行う。
- ▶ 最終的に、すべての添削が完了した後、「タグ化」を行う。

コーパス構築作業：タグ付け

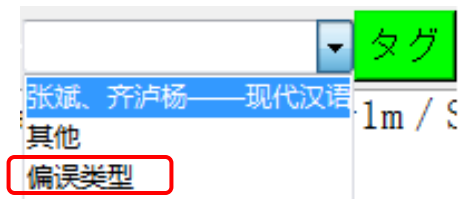
- ▶ TNR_ChineseErrorCorpusTagger3.0_notag.exe を開く。



以下のように作文が読み込まれる。先ほどの添削箇所が表示される。

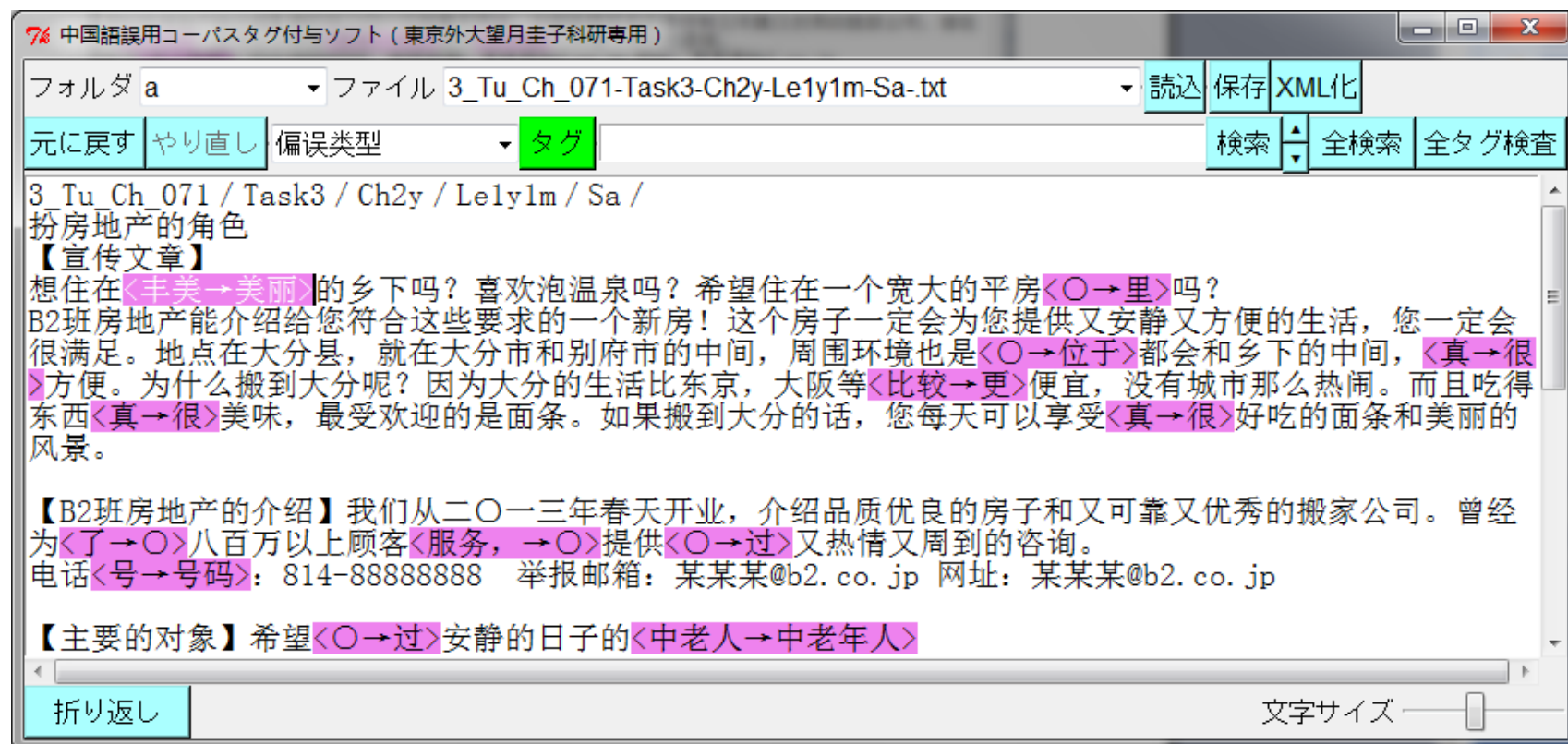


タグ付け「誤用タイプ」_1



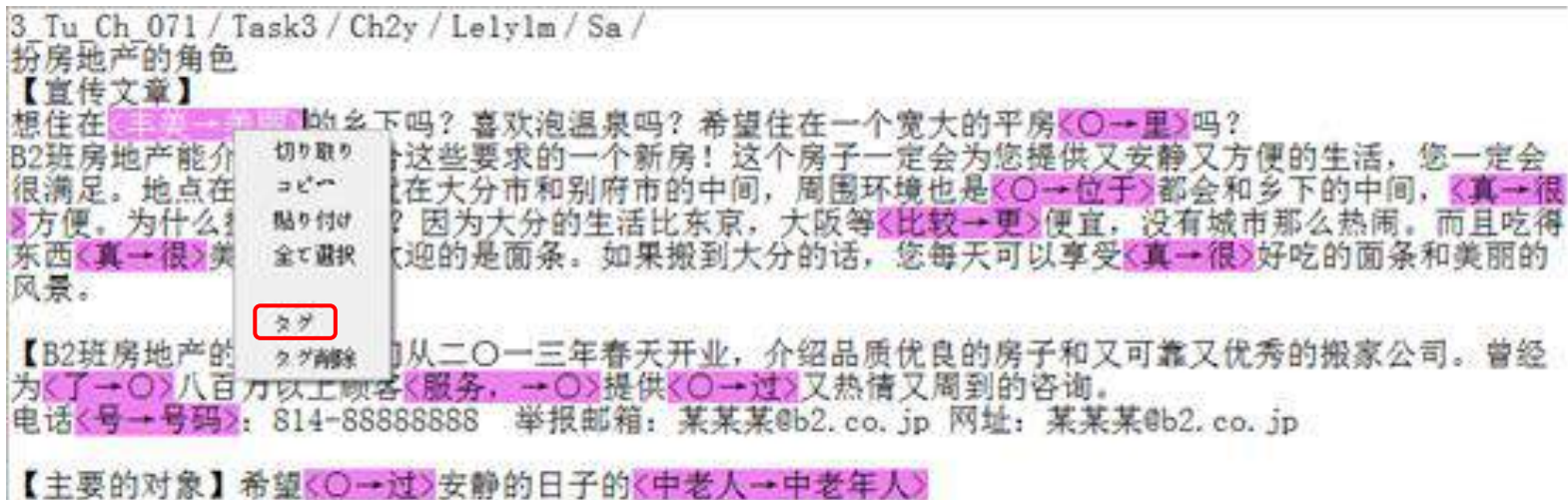
タグリストの中から“偏误类型”を選択する。

タグを付けたい箇所（〈丰美→美丽〉）をドラッグする。



タグ付け「誤用タイプ」_2

右クリックして、リストの中から「タグ」を選択する。



その結果、以下のような画面が表示され、「替換」をクリックして、「タグ付与」をクリックする。



タグ付け「誤用タイプ」_3

結果、タグとして「替換」が表示される。

7% 中国語誤用コーパスタグ付与ソフト (東京外大望月圭子科研専用)

フォルダ a ファイル 3_Tu_Ch_071-Task3-Ch2y-Le1y1m-Sa-.txt 読込 保存 XML化

元に戻す やり直し 偏誤タイプ タグ 検索 全検索 全タグ検査

3_Tu_Ch_071 / Task3 / Ch2y / Le1y1m / Sa /

扮房地产的角色

【宣传文章】

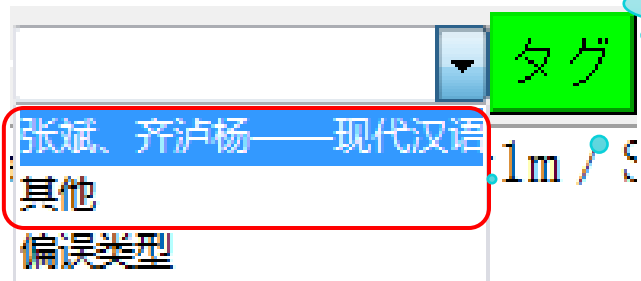
想住在<替换 / 丰美→美丽>的乡下吗？喜欢泡温泉吗？希望住在一个宽大的平房<○→里>吗？B2班房地产能介绍给您符合这些要求的一个新房！这个房子一定会为您提供又安静又方便的生活，您一定会很满足。地点在大分县，就在大分市和别府市的中间，周围环境也是<○→位于>都会和乡下的中间，<真→很>方便。为什么搬到大分呢？因为大分的生活比东京，大阪等<比较→更>便宜，没有城市那么热闹。而且吃得东西<真→很>美味，最受欢迎的是面条。如果搬到大分的话，您每天可以享受<真→很>好吃的面条和美丽的风景。

【B2班房地产的介绍】我们从二〇一三年春天开业，介绍品质优良的房子和又可靠又优秀的搬家公司。曾经为<了→○>八百万以上顾客<服务，→○>提供<○→过>又热情又周到的咨询。电话<号→号码>：814-88888888 举报邮箱：某某某@b2.co.jp 网址：某某某@b2.co.jp

【主要的对象】希望<○→过>安静的日子的<中老人→中老年人>

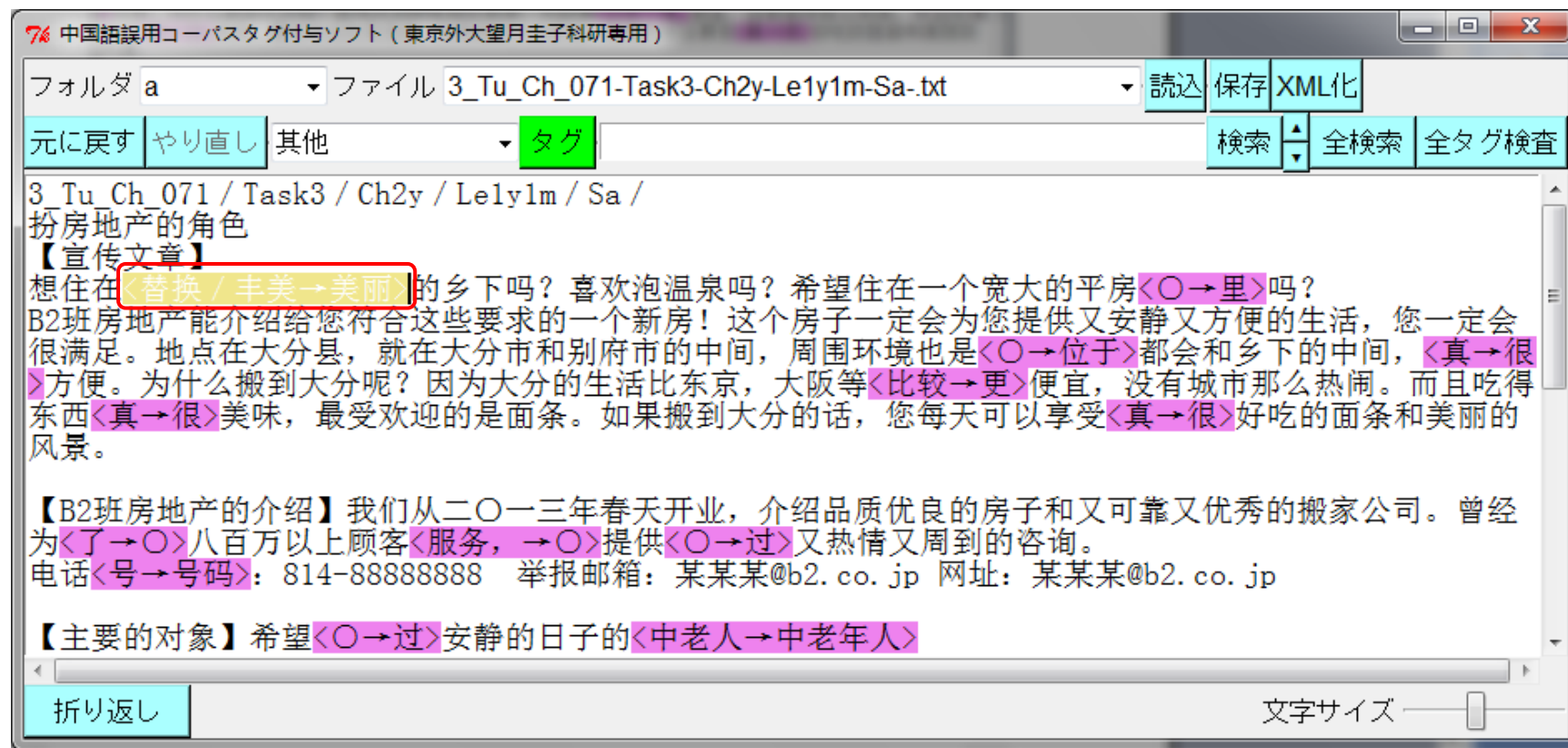
折り返し 文字サイズ

タグ付け「文法カテゴリー」_1



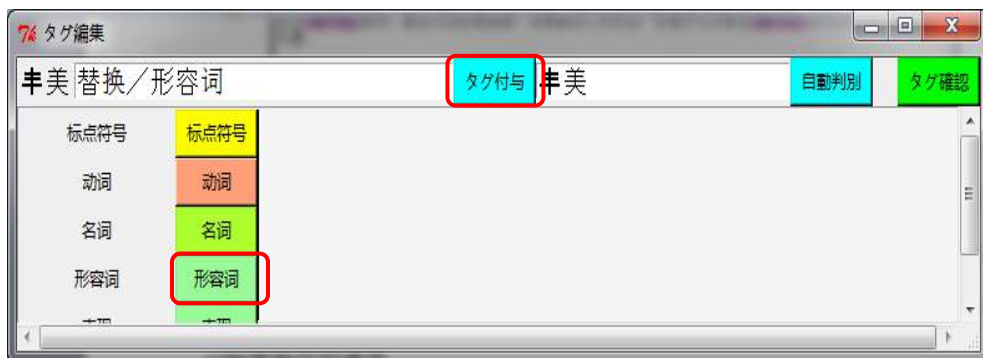
“其他”は、“张斌, 齐沪杨”になかった項目を付け加えたものである。

タグリストの中から“张斌, 齐沪杨”或いは“其他”を選択する。



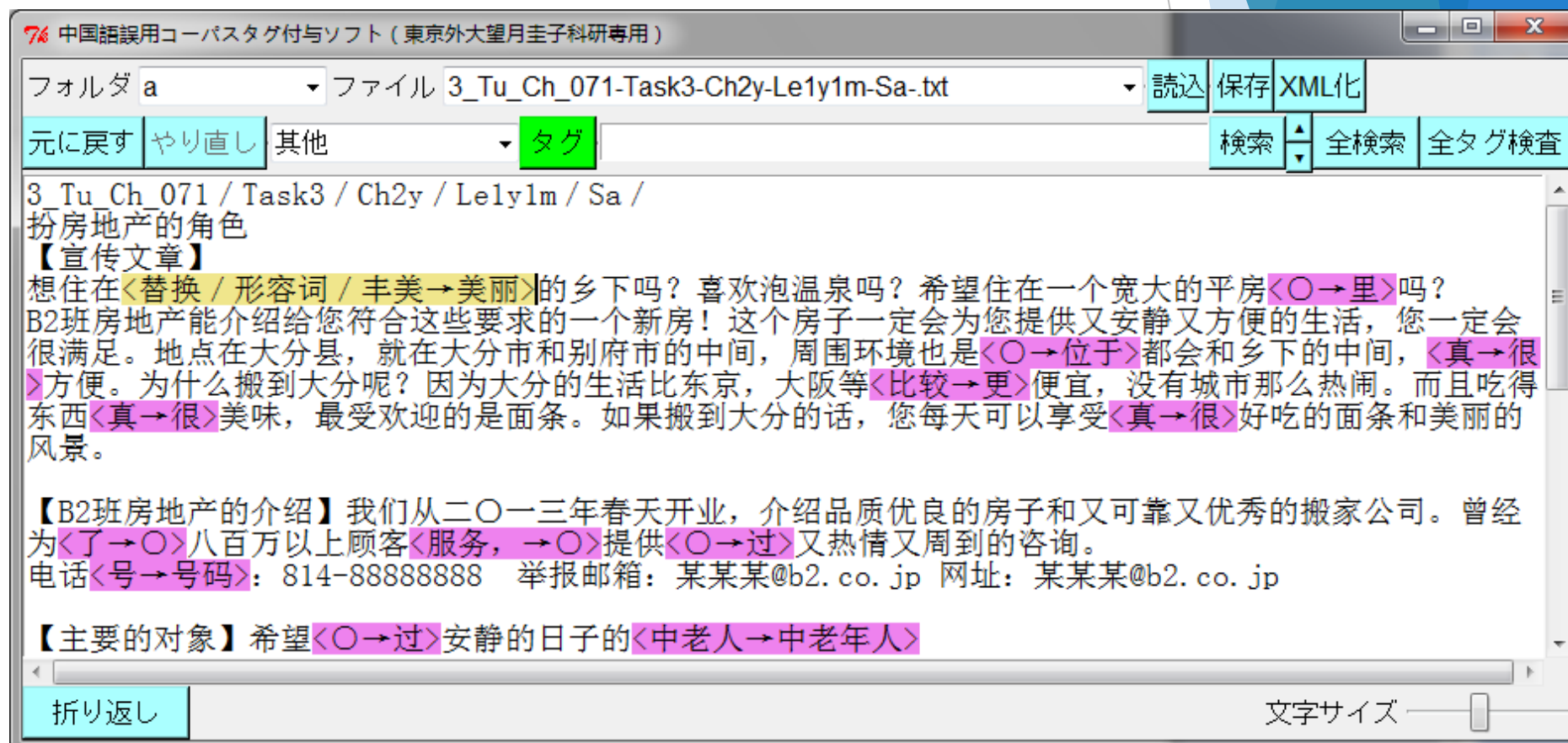
タグを付けたい箇所
（<替换 / 丰美 -> 美丽>）をドラッグする。

タグ付け「文法カテゴリー」_2



この場合は、「形容词」をクリックして、「タグ付与」をクリックする。

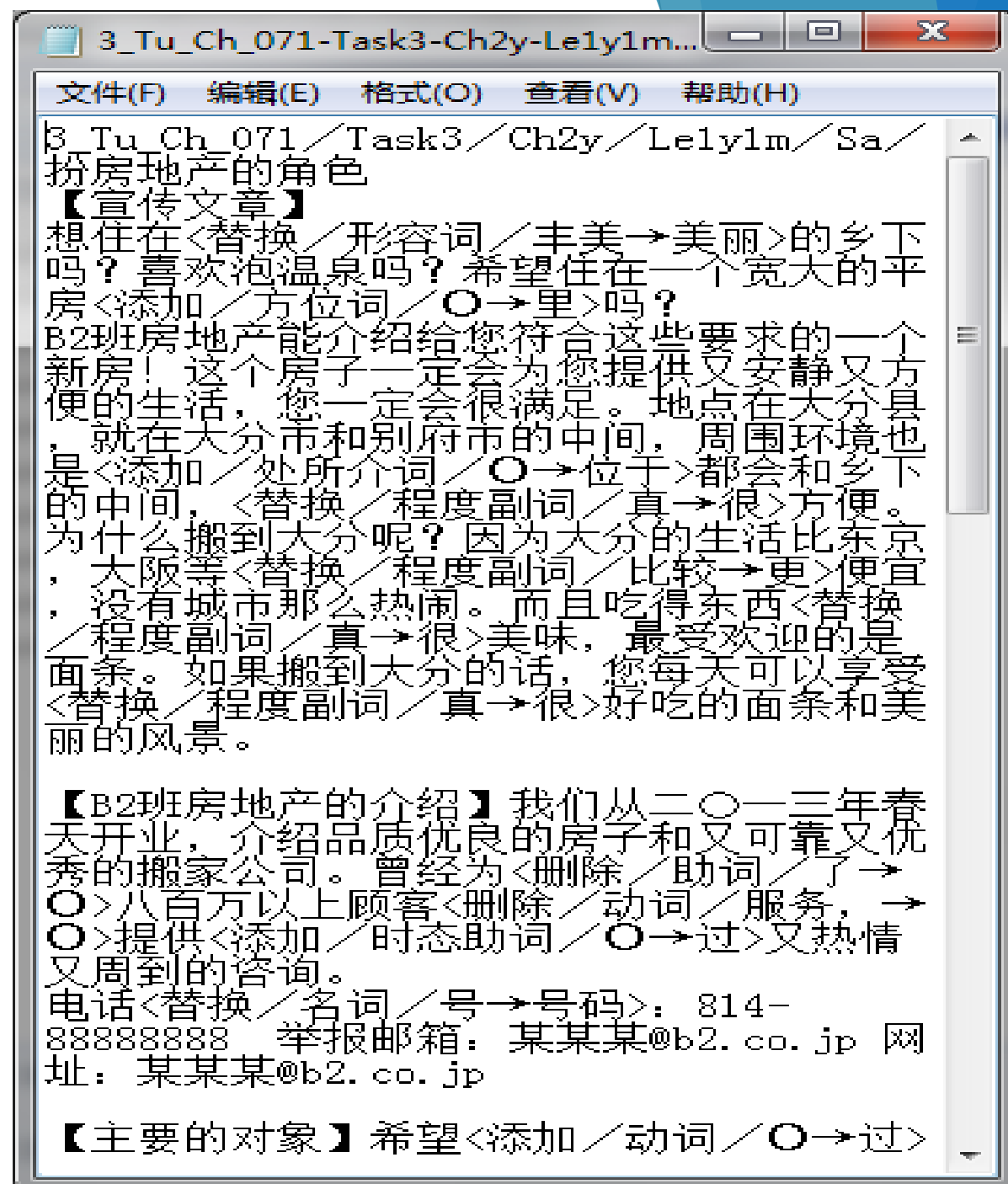
結果、タグとして「形容词」が表示される。



XML化

読込 保存 XML化

- ▶ 「XML化」をクリックする。
- ▶ 最終的に作成したXMLファイルをもとにして、検索エンジンを構築する。



ご清聴ありがとうございます。

2014. 5. 29