

# 翻译软件 V1.30 版升级说明

## V1.30 版主要升级内容：

1. 加入了**字数统计**和**翻译进度**信息的支持。
2. 增加了“**相似度分析**”功能，用于分析待译文档内的句子相似程度，也可以将待译文档与已经翻译过的同类型文档进行比较，统计出完全相同的句子、50%以上相同的句子、不相同的句子，标准版可以将这些句子导出以便做进一步的处理。
3. 增加了“**文档分析**”功能，用于分析未翻译句子在记忆库中的匹配程度，通过与记忆库中例句的比较，统计出完全匹配的句子、50%以上匹配、不匹配的句子，标准版可以将分析的结果导出做进一步的处理。
4. 增加了“**自动短语提取**”功能，通过对原文、译文的分析，可以从已经翻译的文档、记忆库、双语对齐项目中自动提取短语（同时提取短语的原文及译文）。
5. 标准版增加了“**文件拖放拷贝**”功能，同时打开多个项目文件窗口，通过文件拖放操作，可以方便地实现项目文件之间的快速复制。方便对项目文件中翻译文档的整理及双语对齐项目中记忆库的分类整理。
6. **对记忆库中提示的句子直接修改**。记忆库搜索和片断搜索时增加了右键菜单，支持了修改句子、删除句子、定语句子和核准服务器中的句子。如果句子来自于项目文件并且备注为空时，则在记忆库和片断搜索窗口的备注列中显示文件名称；如果句子来自于服务器，则在备注中显示上传人的名称。还支持了在客户端软件中直接修改、删除和核准服务器中的句子。
7. 服务器设置增加了“**允许使用未审核术语**”的选项，选项在重新连接服务器后生效。
8. 标准版增加了**上传已审核术语库**的支持。
9. 标准版增加了**上传已审核记忆库**的支持。
10. **正则表达式词典**的支持，可以使用正则表达式定义翻译的短语或者规则词典的类别。
11. 规则词典加入的查找记录的支持。
12. 片断搜索加入**按钮“C”**，用于设置搜索时严格匹配鼠标划选的词语，搜索结果不受大小写变化、词语的变化形式等因素的干扰，对缩略语等的搜索比较适用。
13. 项目文件支持了设置句子标志的支持，用于统一删除句子的标志及核准所有未翻译的句子。
14. 导出编译文本时，如果句子译文为空并且标注为正确时导出句子原文。
15. 项目文件增加了清除译文的支持。

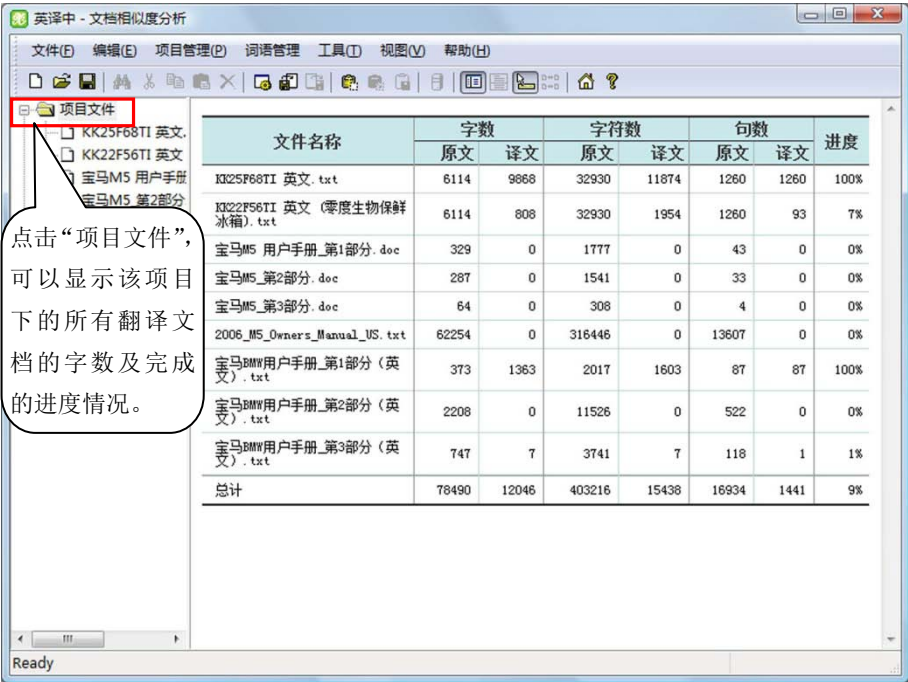
- 16. 查找和替换中的历史记录保存了替换的词语。
- 17. 当译文为空并且原文未选择时, 按[SRC]按钮时复制原文。
- 18. 调整了文件格式, 新版本软件保存的文件不能在旧版本软件中打开。
- 19. 更正了某些情况下完全匹配的句子没有自动填写的问题。
- 20. 更正了导入双语文本格式时某些情况下不能正确对齐的问题。

雪人 CAT 网络协同平台\_服务器版也做了相应的更新, 版本升级到 V1.02。

以下将详细介绍.....

一、字数统计及翻译进度

字数统计及翻译进度统计是一项很实用的功能, 软件将原文和译文的字数、字符数、句子数全部统计出来, 并以表格的方式清晰显示, 如下图。点击屏幕左侧的“项目文件”, 就可以显示该项目下的所有文档的字数及进度完成情况。



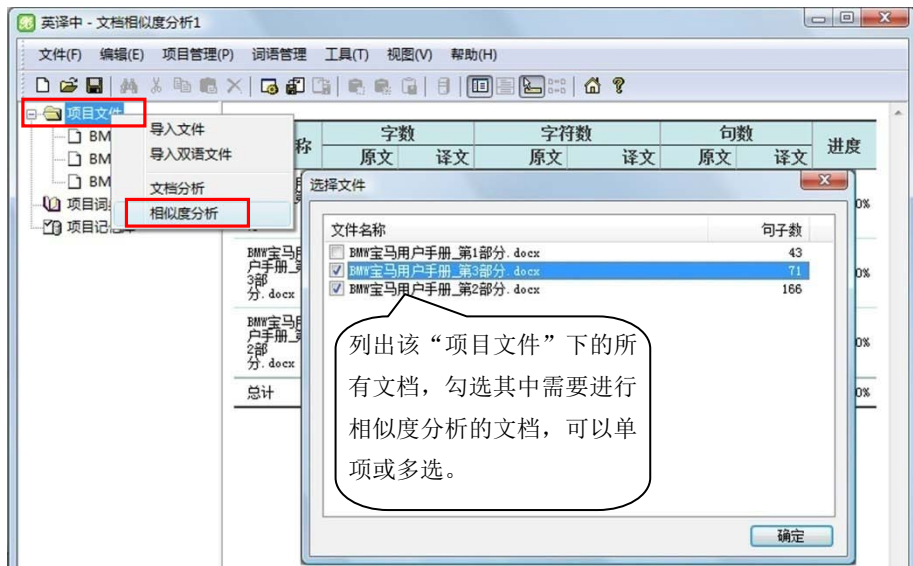
字数统计功能同样适用于“双语对齐”项目, 点击屏幕左侧的“项目文件”, 就可以

显示该双语对齐项目下的所有对齐文档原文、译文、字数、字符数、句对，清楚了解所建句库的规模。



二、相似度分析

**相似度分析**功能用于分析待译文档内的句子相似程度，也可以将待译文档与已经翻译过的同类型文档进行比较，统计出完全相同的句子、50%以上相同的句子、不相同的句子，并可将这些句子按需导出以便做进一步的处理。右键点击屏幕左侧的【项目文件→相似度分析】功能，在弹出的“选择文件”窗口中，会列出该“项目文件”下的所有文档，勾选其中需要进行相似度分析的文档，可以单项或多选，如下图：如果仅勾选一个文档，则仅在该文档中分析其句子的相似情况；如果勾选多个文档，则可以合并分析这些文档中句子的相似程度，例如：如果要翻译的文档与自己曾经翻译的文档类型相似，就可以将它们放在同一个翻译项目下，然后进行“相似度分析”，软件就可以马上列出这些文档之间存在的相似句子。



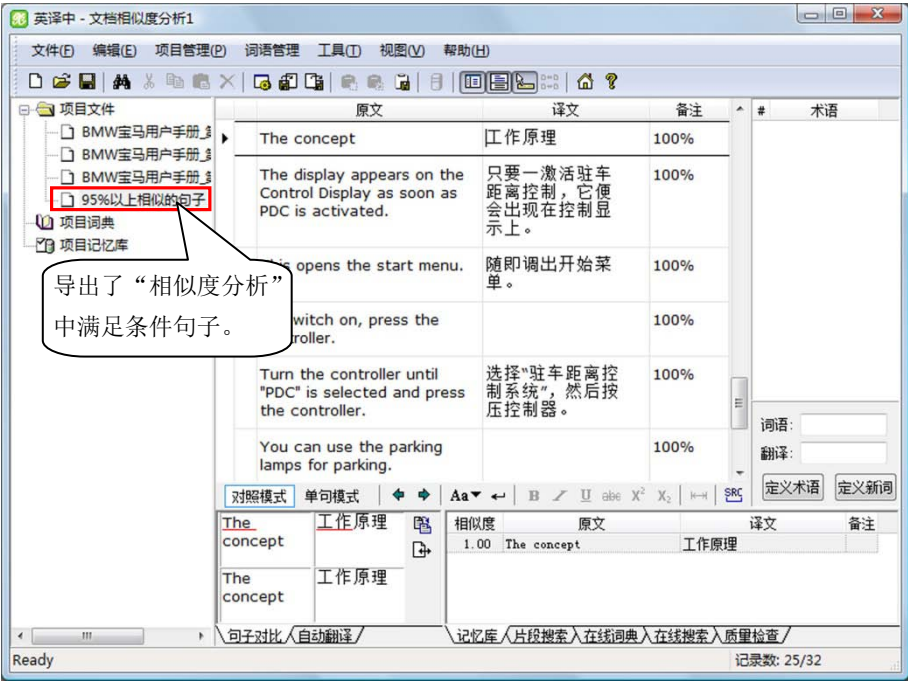
勾选需要进行分析的文档后，点击“确定”按钮，软件会按百分比显示相似的句子数及不匹配的句子数量，如下图：



勾选各个匹配率区间，可以将处于这个匹配率的具体句子导出，例如：上图中勾选了100%匹配和95%~99%区间匹配的句子，然后点击“导出”按钮，在弹出的窗口中输入文件名，例如命名为：“95%以上相似的句子”。如下图：



点击“确认”按钮后，导出的文件会放在左侧的“项目文件”下面，文件名则是上图中输入的文件名，即“95%以上相似的句子”，如下图：

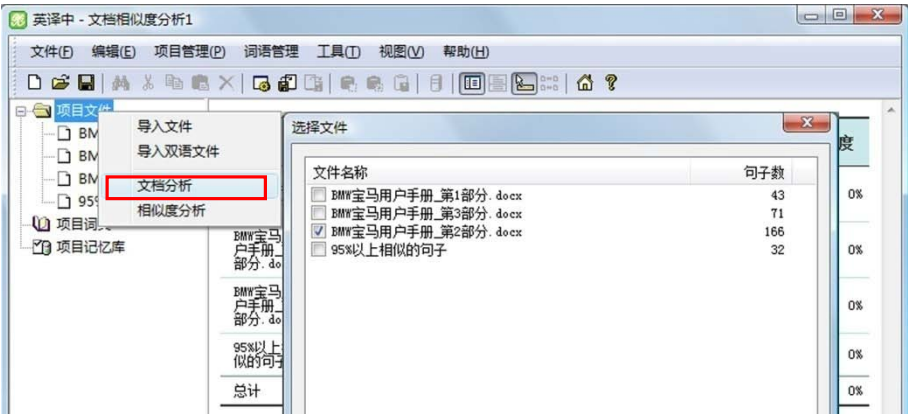


“相似度分析”的结果导出后，我们可以对这些句子进行预先处理，例如先翻译这些相似度高的句子。翻译后它们会被自动保存在项目记忆库中，然后就可以利用记忆库来快速预翻译该项目下的其他文档了；与记忆库中完全匹配的，预翻译时就已经填写了译文，与记忆库中相似的，当翻译该句时就会自动提示。

三、文档分析

“文档分析”与“相似度分析”从操作方法上相类似，但二者用于比较分析的对象不同。**“相似度分析”是对选择的多个相似文档之间存在的相似句子进行分析**，如果仅选择

一份文档，则就对该份文档内存在的相似句子进行分析。“文档分析”是将选择的文档与记忆库中的例句进行比较，统计出待译文档中有多少句子与记忆库中的例句匹配。



右键点击屏幕左侧的【项目文件→文档分析】功能，在弹出的窗口中勾选需要进行相似度分析的文档，可以单项或多选，如上图：然后点击“确定”按钮，就可看到记忆库中的例句与所选择文档的相匹配。如下图：

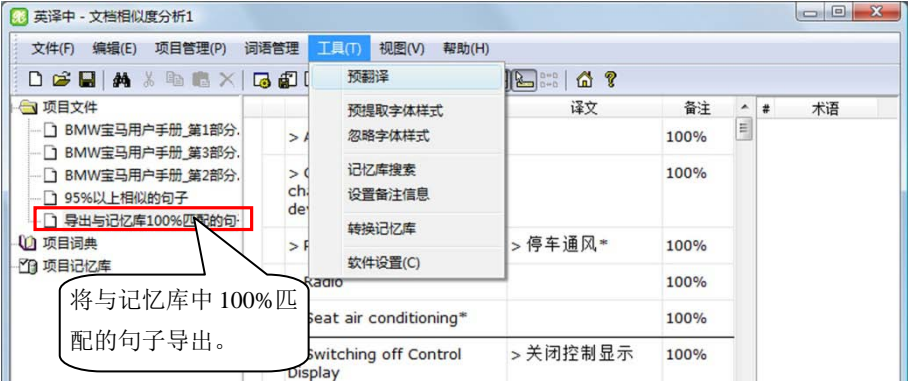


勾选具体的匹配区间，然后点击“导出”按钮，并给导出的文件命名，如下图：导出的文件命名为“导出与记忆库 100%匹配的句子”。





点击“确定”按钮后，导出的句子会放在“项目文件”夹下面，如下图：



除了可以导出与记忆库 100%匹配的句子,也可以导出与记忆库 50%以上相似的句子,这些句子有记忆库为参考,翻译起来会省事很多。还可以将那些与记忆库不匹配的句子导出,例如,下图中“导出与记忆库中不匹配的句子”。

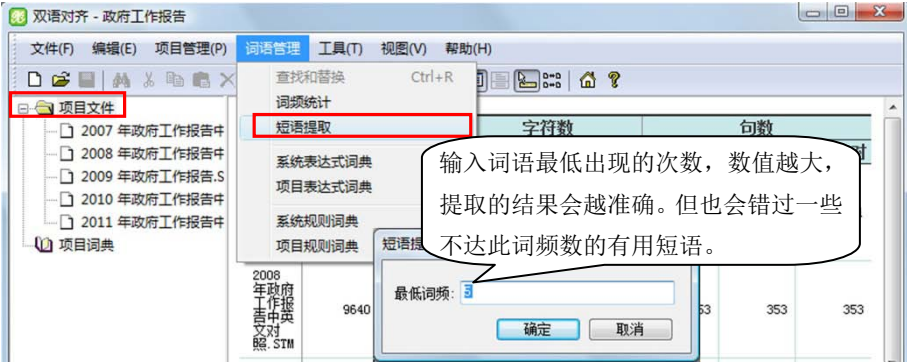


通过“相似度分析”和“文档分析”，译员可以很好地了解待译文档的难易程度，给翻译项目报价和估计完成的时间等方面提供了很好的参考。

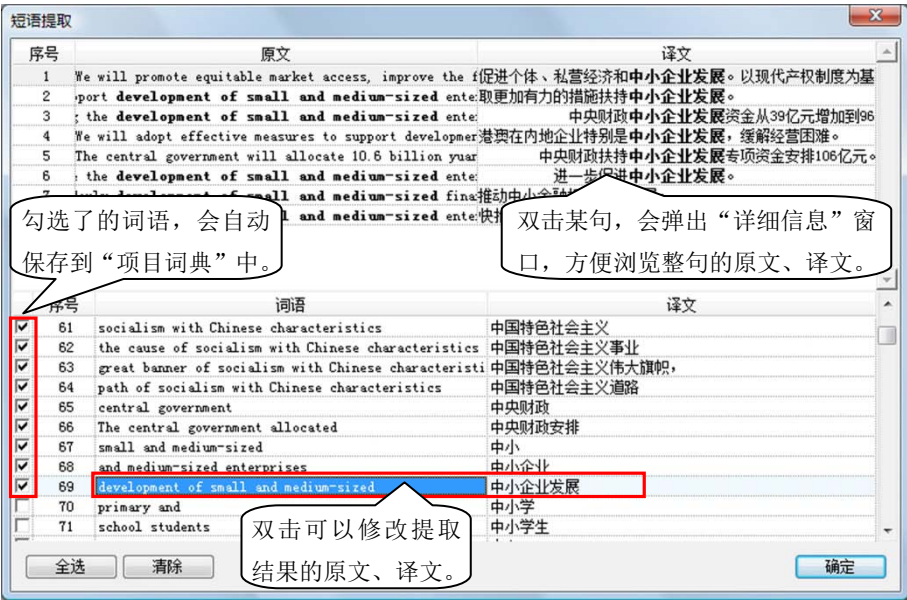
四、自动短语提取

**自动短语提取**功能是一项非常有用的功能，它能从记忆库、双语对齐项目、已经翻译过的文档中通过原文、译文的对应分析提取可能的短语，经审核后就可以加入到用户的“项目词典”中。下面以几份“政府工作报告”的双语对齐项目为例，介绍其操作方法：首先点击屏幕左侧的“项目文件”，然后点击【词语管理→短语提取】菜单（注意：若点击“项目文件”后点击“短语提取”功能，则提取的范围是该项目文件下的所有文档；若点击某个具体的文档后点击“短语提取”功能，例如下图的“2007 年政府工作报告”，则仅在该

文档中分析、提取)。点击“短语提取”功能后，会弹出窗口要求输入“最低出现的词频数”，输入的数值越大，提取的结果会越准确，但也会错过一些不达此词频数的有用短语。输入的数值越小提取出来满足条件的结果会越多，提取时耗费的计算分析时间也会越长。如果是对记忆库进行提取，建议记忆库的数量不要超过 10 万句对之内，太大了计算的时间会比较长并且有可能由于内存不足而异常关闭软件。

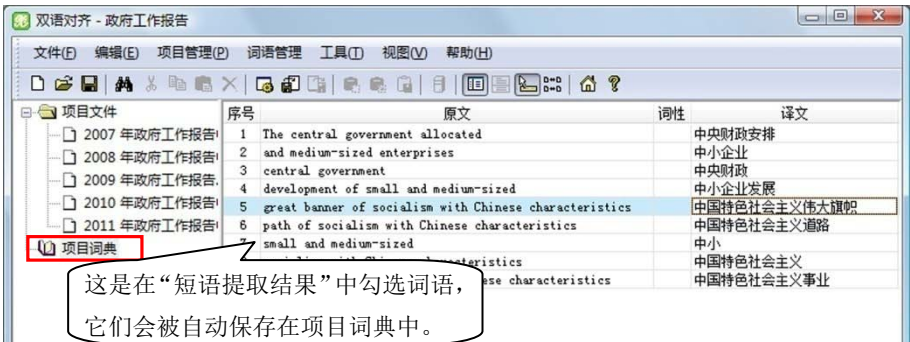


下图是输入最低词频数为 5 时的短语提取结果，对正确的词语打勾后它就会自动保存到“项目词典”中。对需要修改的词语，可以用鼠标双击它进行修改。



上图中，勾选的词语，将会被保存在“项目词典”中，如下图：

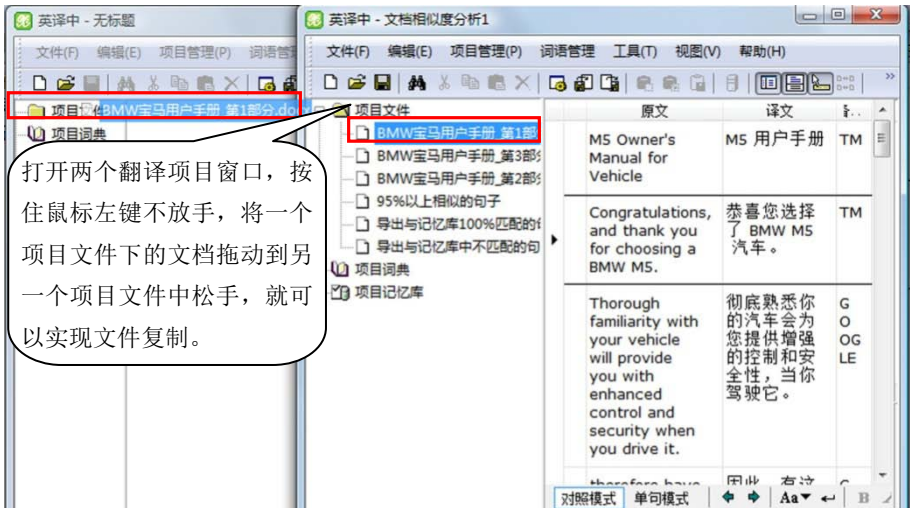




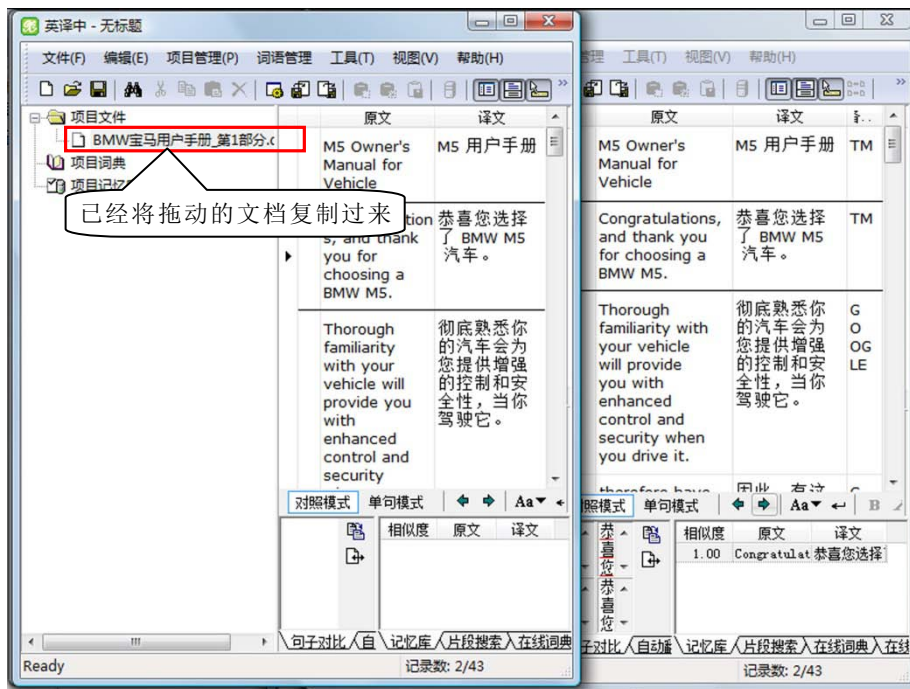
## 五、项目文件的拖放复制

雪人 CAT 中一个“项目文件”下面支持导入 200 份翻译文档，例如将多份需要翻译的文档导入在同一个“项目文件”下面，对它们进行统一的文档分析后，又需要分派不同的人完成，这时方便的做法是：新建几个翻译项目，分别将各份待译文档拖到新的项目文件下面，然后将这些项目文件发给其他人。他们译完后，再将他们翻译后的文档拖回到原来的项目文件中，这样就保证了项目文件的统一，方便存档及日后查看。

文件拖放复制的操作方法：将两个项目文件窗口平铺在桌面，点击需要复制的文档后按住鼠标左键不松手，将它拖到另一个窗口的“项目文件”下面，此时鼠标状态会出现一个“+”号，如下图：



松开左键，拖放的文件就会被复制到新的翻译项目中，如下图：

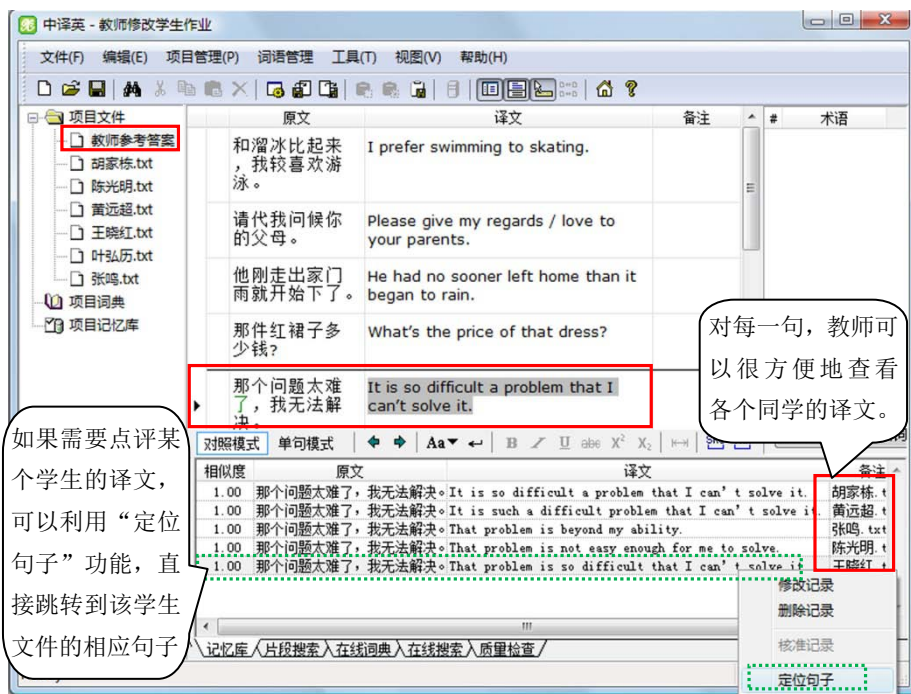


对于普通译员，文件复制功能也是很实用的。例如：自己在一个项目文件下面翻译积累了很多的文档，若需要将其中的某一份文档发给他人审校，自然不希望发整个项目的 STP 文件，但逐个删除其他的文档又比较麻烦。审校完了，发回了的文件又怎么加入到自己原来的翻译项目中，这些问题在 V1.29 版之前都不好解决。在 V1.30 后，利用文件拖放复制功能，将需要审校的文件拖出后发给他人审校，审校完了，在将它拖入原来的项目文件就可以了。

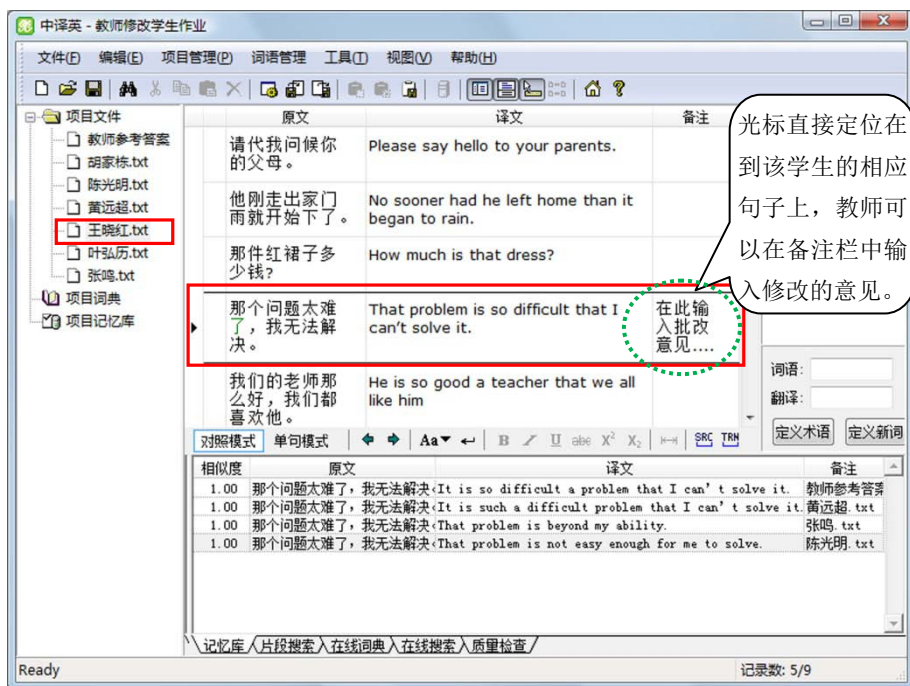
文件拖放功能，在“双语对齐”项目中也很有用，例如：之前整理对齐的文章可能比较多、比较杂，没有进行系统的分类。现在就可以新建几个双语对齐项目，然后将之前对齐的各种文章，通过拖放复制的操作进行分类整理，这样就可以分类导出记忆库。

下面讲述“文件拖放复制”功能对翻译教师修改学生作业方面的应用。用雪人 CAT 的免费版进行翻译教学，是个不错的选择。教师布置的翻译作业以文本方式发布，这样学生只要下载雪人 CAT 的免费版就可以进行翻译了。学生将翻译好的 STP 项目文件发送给老师，老师修改作业时，首先自己新建一个翻译项目，并做好参考译文，然后可以将学生们的 STP 文件全部拖到这个项目文件，这样做的好处是：当鼠标点击每一句时，记忆库中就会自动显示来自各个同学的译文，方便教师集中发现学生翻译时存在的问题。如下图：当教师修改“那个问题太难了，我无法解决。”这一句时，在屏幕的下方就自动提示出多个学生的

译文。



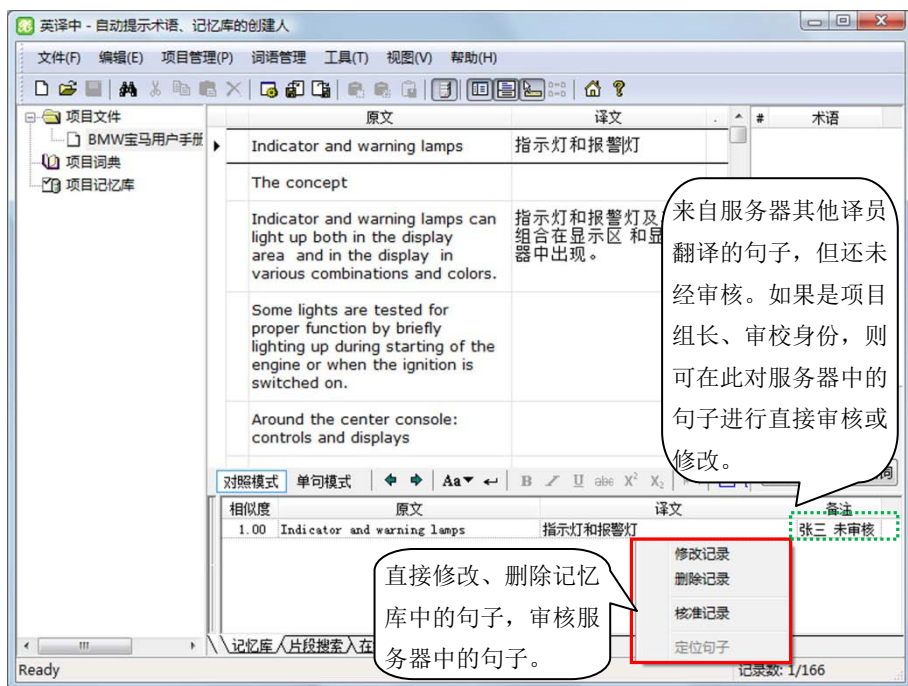
如果需要点评某个学生的译文, 鼠标点击该句后 (如上图的学生“王晓红”的译文), 再点击【右键菜单→句子定位】功能, 此时光标就会直接跳转到该学生 (即“王晓红”) 文件的相应句子上面, 如下图:



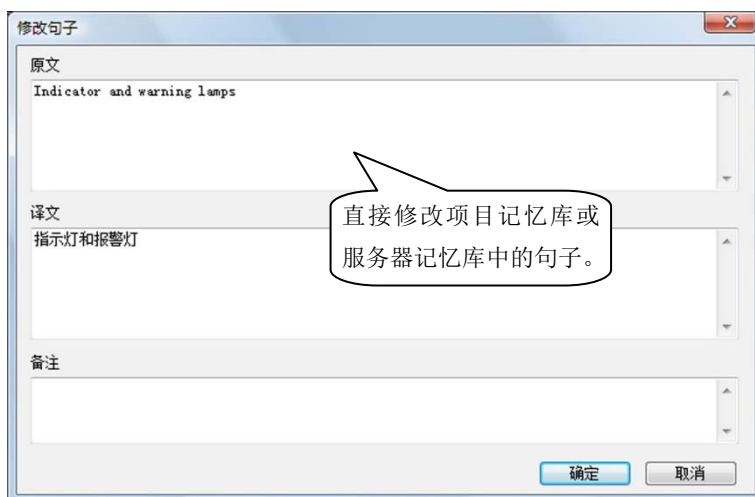
教师可以在“备注”栏中填写批改意见并使用[F4]或[F9]键设置标志。这样集中批改完后，可以将整个项目文件发给学生，学生不但可以看到自己译文及老师的批改，也可以看到其他同学的译文及老师的批改意见，可以很好地促进同学间的互相学习，相信教学效果会更好。

## 六、记忆库的修改、核准、上传

在翻译时有时会发现屏幕下方的记忆库或片段搜索窗口中提示出来的句子有错，想顺便修改错误，但是回到“项目记忆库”中查找出错误的句子通常比较麻烦。在 V1.30 后这个问题就好解决了，V1.30 在记忆库中记忆库搜索和片段搜索窗口中增加了右键菜单，支持了修改句子、删除句子、定语句子和核准服务器中的句子，可以很方便地对记忆库中提示的句子进行删除、修改等操作，如下图：



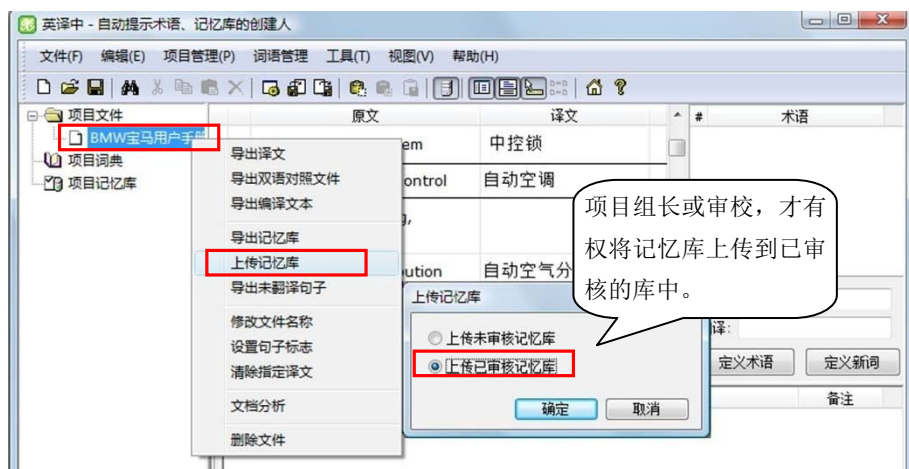
在记忆库或片段搜索窗口中，点击【右键菜单→修改记录】功能，可以修改所提示的记忆库中的句子。如下图：注意：对项目记忆库、服务器中的记忆库中的句子，可以直接进行修改、删除操作，如果记忆库是通过“添加”引用的外部记忆库文件，则不能进行此操作。





如果句子来自于项目文件并且备注为空时，则在记忆库和片断搜索窗口的备注列中显示文件名称（例如：前面在讲述“教师批改学生作业”案例时，记忆库中自动备注的学生姓名）；如果句子来自于服务器，则在备注中显示上传人的名称，如上页图中“备注”显示的内容：“张三 未审核”，如果当前用户是项目组长或审校的身份，则就可以在此对服务器中其他译员翻译的句子直接修改；点击【右键菜单→核准记录】功能则可以直接审核通过。

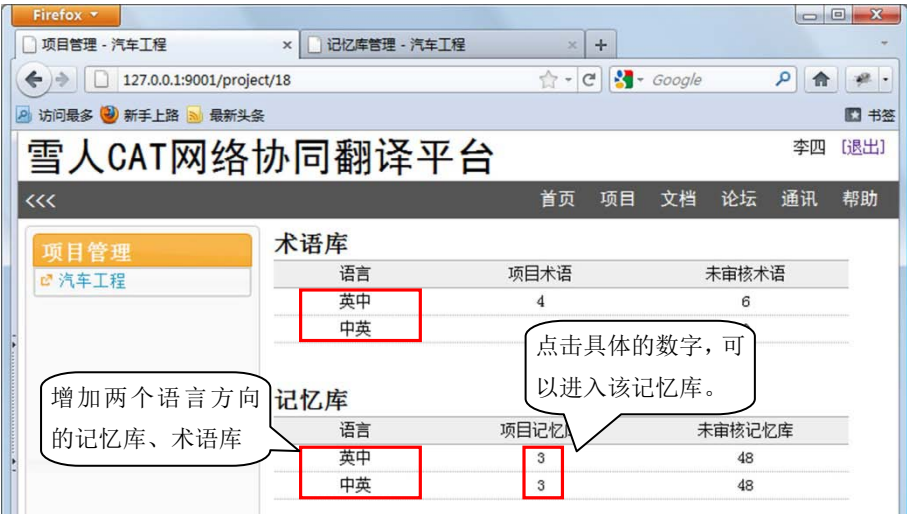
虽然在记忆库和片段搜索窗口中，可以利用【右键菜单→核准记录】功能来对其他译员翻译的句子进行审核，但这种审核方式效率不高。而登录服务器进行审核，又脱离了原文的段落衔接等整体篇章信息。所以，最高效的审校工作模式应该是：审校将项目组中各成员要翻译的文档全部导入到一个雪人 CAT 中的项目文件，其他译员翻译的句子会自动同步到服务器，审校立即可以在客户端界面中看到他们翻译的句子，正确的直接上屏，不正确的则修改，这样在审稿时就可以很方便地把握原文的整体性。审核完后，点击屏幕左侧的翻译文件，再点击【右键菜单→上传记忆库】功能，在弹出的窗口中选择“上传已审核记忆库”选项，这样就可以集中上传审核后记忆库。



## 七、V1.02 版服务器的更新

在 V1.02 的服务器版本中也有内容更新，如下图：服务器的记忆库和术语库都划分了两个语言方向，如中英、英中。二者内容是一样的，只是原文和译文的前后顺序显示不同。





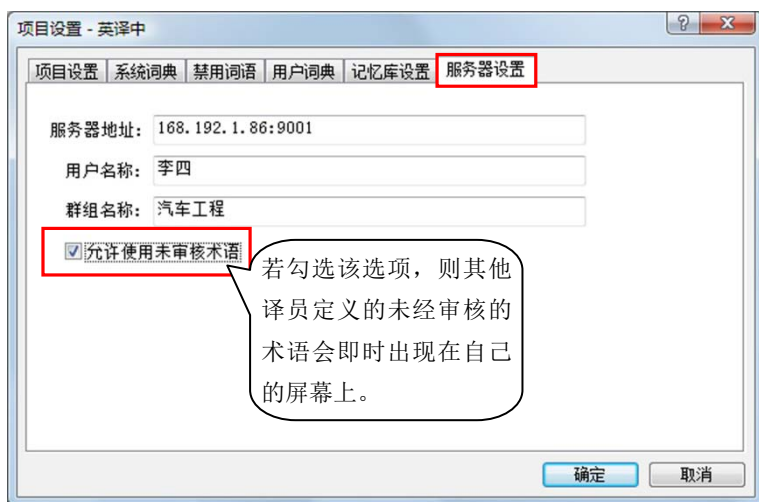
点击上图中具体的数字，可以进入到该记忆库或术语库。如下图：在记忆库或术语库中，点击具体的译员名字，可以筛选出来在该译员的句子或术语。另外，在 V1.02 中还更改了页数的显示方式，可以直接跳转到目标页面。




## 八、术语的审核和来源提示

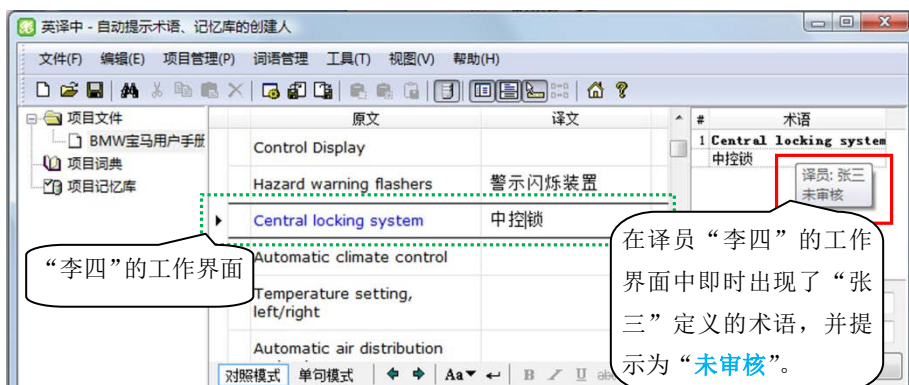
一般情况下，服务器中的术语是应该由项目组长或审校审核后，同项目组的其他译员

才能使用。但雪人 CAT 也对术语的使用提供了一个灵活的方式，在【项目管理→项目设置→服务器设置】中，增加了一个“**允许使用未经审核术语**”选项，如下图：用户可以按实际需要决定自己当前翻译的项目是否勾选这个选项。若勾选该选项，则其他译员定义的未经审核的术语会即时出现在自己的屏幕上。



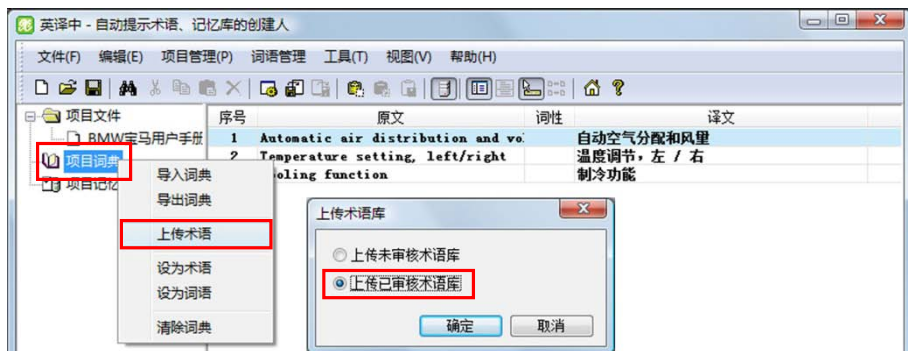
注意：勾选“允许使用未经审核术语”选项后，需要点击工具栏的  “服务器连接图标”将服务器断开，然后再点击该图标链接上服务器才会生效。

如下图中，在译员“李四”的工作界面中，即时出现了“张三”定义的术语，并提示为“**未审核**”（鼠标指向屏幕右侧的术语时，会自动弹出该术语的创建人、审核与否等信息）。

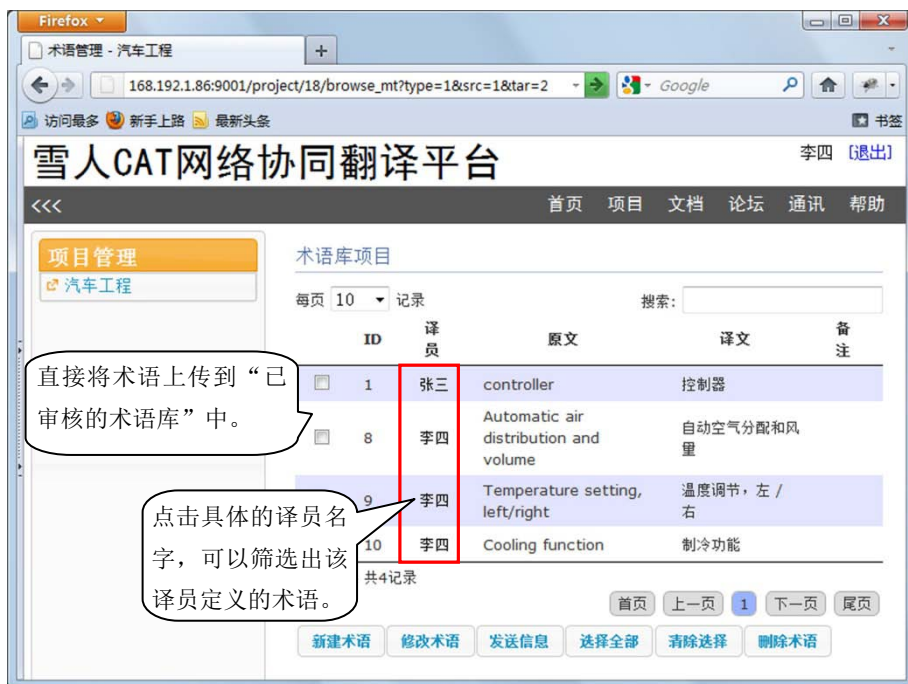


如果翻译时由于某种原因没有链接服务器，或者有些现成的术语资源需要传到服务器中，则可以通过“上传术语”功能将它上传到服务器。点击屏幕左侧的“项目词典”，然

后点击【右键菜单→上传术语】功能，如果是“项目组长或审校”身份，此时会弹出一个窗口询问：将术语上传到未审核术语库中，还是直接上传到已经审核的术语库中；普通用户身份则不会弹出询问窗口，而是直接上传到“未审核的术语库”中，如下图：

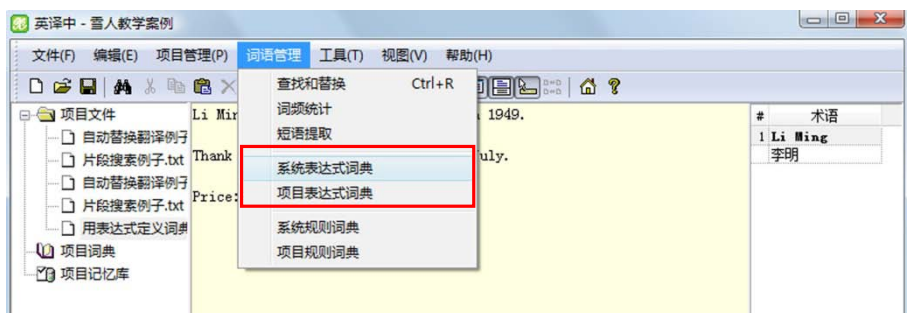


如果确定上传的术语正确无误，则可以选择“上传已审核术语库”，这样就不用再到服务器中进行术语审核工作了，如下图，所上传的术语直接进入审核的术语库中。

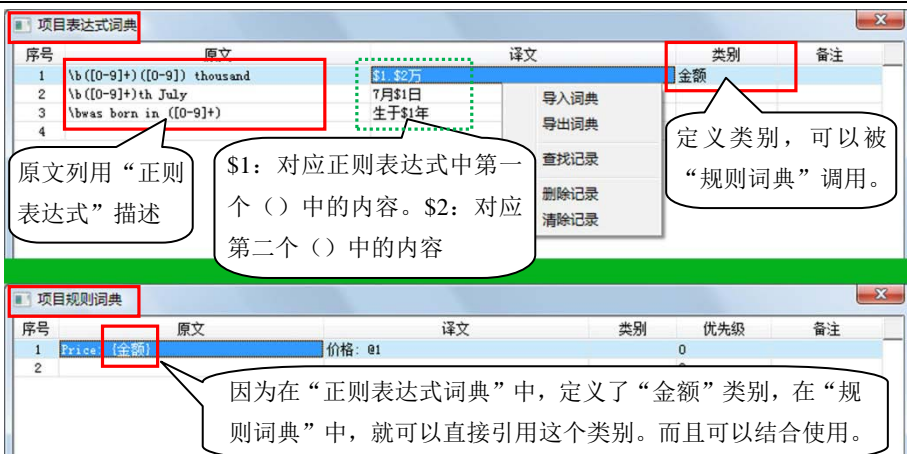


## 九、用正则表达式定义词典

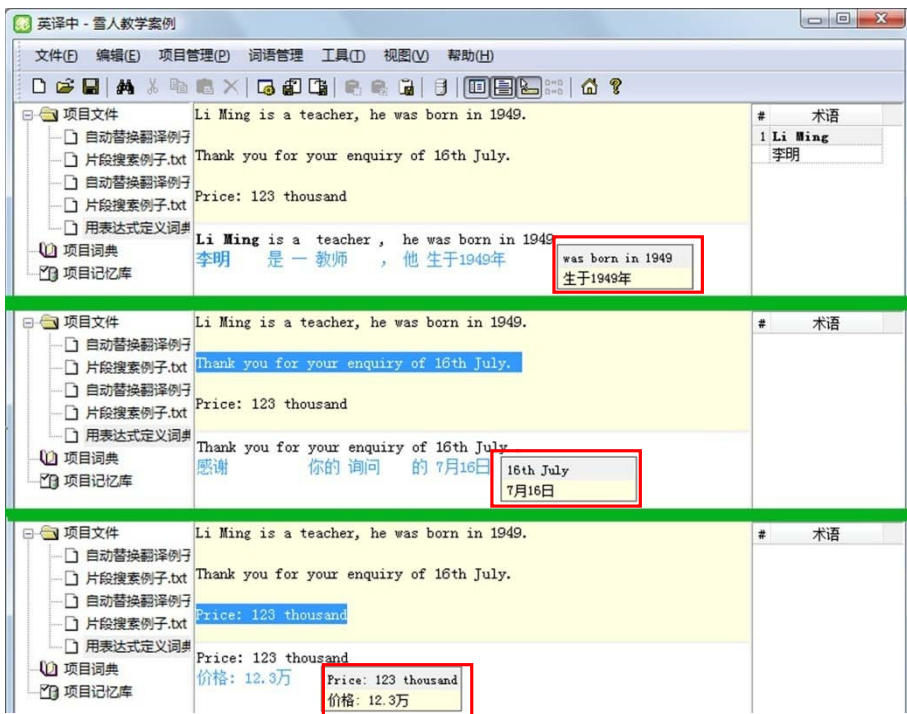
在 V1.30 中增加了**正则表达式词典**的支持，可以使用正则表达式定义翻译的短语或者规则词典的类别。熟悉正则表达式使用的用户会感受到这项功能带来帮助的。点击【**词语管理**】菜单，其下面增加了【**系统表达式词典**】和【**项目表达式词典**】这两个菜单，如下图：如果在【**系统表达式词典**】中进行了定义，则所有的翻译项目都可以使用；如果在【**项目表达式词典**】中进行定义，则仅当前翻译项目可以使用，其他项目需要使用时，可以通过【**右键菜单**→**导出词典**→**导入词典**】来使用。【**表达式词典**】既可以单独使用，也可以通过“**类别**”定义与【**规则词典**】**联合使用**。（注：正则表达式词典比较消耗计算资源，一般情况下建议不要超过 1000 条规则）



例如我们希望对“Li Ming is a teacher, he was born in 1949. Thank you for your enquiry of 16th July. Price: 123 thousand”这几句的日期、金额等自动正确翻译。则可以在“项目表达式词典”中做如下图的定义：通过定义“\bwas born in ([0-9]+) 生于\$1年”，就可以将这个格式的句子自动翻译成“生于多少年”；通过定义“\b([0-9]+)th July 7月\$1日”就可以将所有这种格式的翻译成“7月多少日”。通过定义“\b([0-9]+)([0-9]) thousand \$1.\$2万”就可以将英语中的多少十千、百千等情况翻译成汉语的多少万，例如 56thousand 译成 5.6 万，其中 \$1 对应的是正则表达式中第一个( )中的内容，即([0-9]+)，\$2 对应的是正则表达式中第一个( )中的内容，即([0-9])。其他替换符号还有 \$0 或者 \$& 表示匹配的全部内容，\$\$ 表示“\$”符号。另外，我们还在这个规则的“类别”列中输入了“金额”两个字，其目的是为了这个正则表达式的内容能被“项目词典”调用。将“正则表达式词典”和“规则词典”连用，可以发挥更加强大的作用。



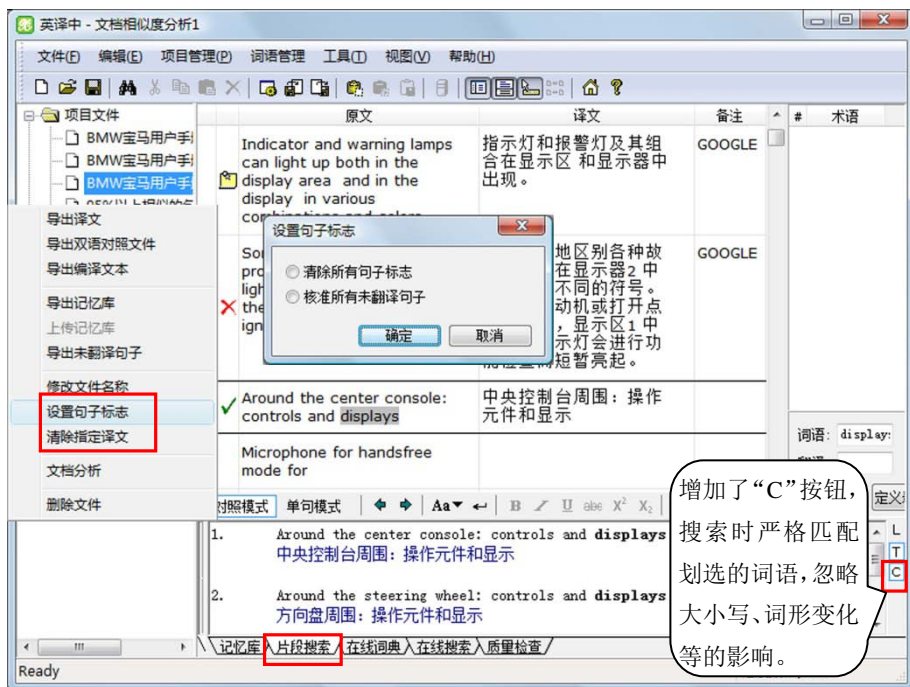
例如上图中，在规则词典中定义“Price: {金额} 价格: @1”，这个“{金额}”类的调用就是指“正则表达式词典”中的“\b([0-9]+)([0-9]) thousand \$1.\$2 万”这个规则。日后，当遇到“Price:123 thousand”这样的短语时，屏幕取词窗口就会将它们作为一个整体，翻译成“价格 12.3 万”，如下图：诸如此类，“Price:4567 thousand”就会翻译成“价格: 456.7 万”。





## 十、其他更新内容

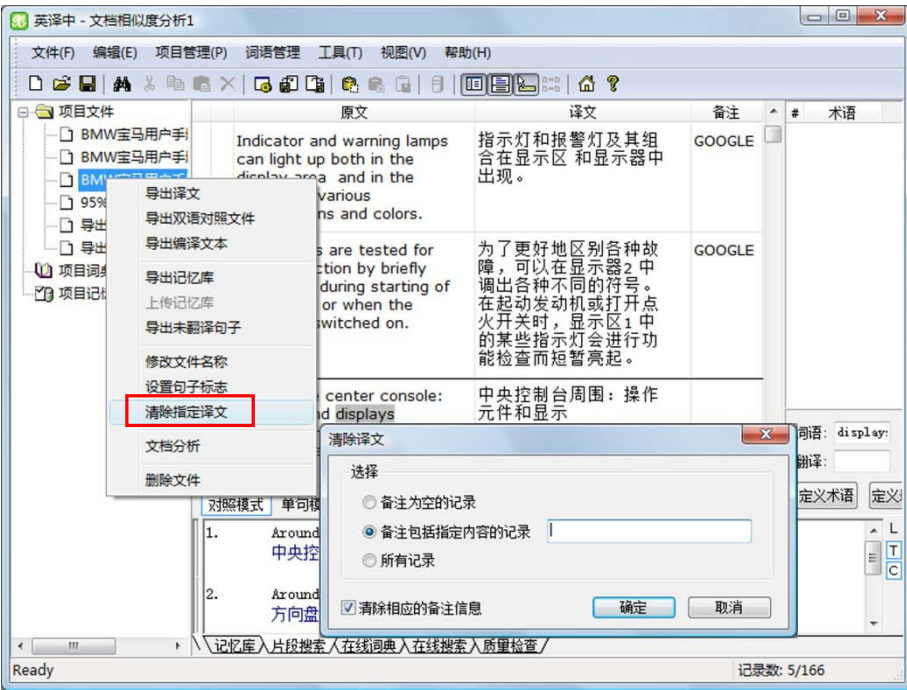
在 V1.30 中，右键点击“项目文件”下面的翻译文档，在右键菜单中增加了【设置句子标志】和【清除指定译文】两个功能，其中【设置句子标志】有“清除所有句子标志”和“核准所有未翻译句子”两个选项，如下图：



“清除所有句子标志”可以将通过【F9】或【F4】标注的句子正确与否的标志全部去掉。“核准所有未翻译句子”是将所有未翻译的句子都标注为正确的句子，相当于逐个按【F9】键的作用，这样在“质量检查”就不会再提示这些未翻译的句子。

【清除指定译文】功能是以备注信息为筛选条件，清除指定的译文。如下图：例如，有时利用 GOOGLE 进行了全文预翻译后，感觉质量不够好，需要将译文全部删除，此时，就可以在“清除译文”的窗口中，输入备注信息为“GOOGLE”，就可以一次性删除所有来自 GOOGLE 的译文。





在规则词典中，增加了“查找记录”的右键菜单，可以通过原文、译文、类别、备注等查找条件，查找所定义的规则。如下图：

