

词义标注语料说明

使用方式：有偿使用。

功 能：可以作为分词、词性标注、词义消歧研究的标准训练语料。

语料来源：人民日报 1998 年上半年的电子版。**语料规模：**从“北大-富士通”语料中随机抽取 11331 句，约 35 万词语。

词 典：《哈工大信息检索研究室同义词词林扩展版》。词典的结构、特点和功能参照“《哈工大信息检索研究室同义词词林扩展版》说明”。

词义编码：采用《同义词词林》的小类代码（第三级代码）。

例外情况：《哈工大信息检索研究室同义词词林扩展版》没有收录的词语的词义一概标成“-1”。

标注示例：

我们/r/Aa02 伟大/a/Ed20 祖国/n/Di02 在/p/Kb01 新/a/Eb28 的/u/Kd01 一/m/Dn04 年/q/Ca18 ， /wp/-1 将/d/Ka13 是/v/Ja01 充满/v/Jd06 生机/n/Dd14 、 /wp/-1 充满/v/Jd06 希望/n/Df08 的/u/Kd01 一/m/Dn04 年/q/Ca18 。 /wp/-1

又/d/Ka14 好像/v/Jb03 倘佯/v/Fb01 在/p/Kb01 自然美/n/Eb30 的/u/Kd01 世界/n/Di01 ， /wp/-1 感受/v/Gb02 乡情/n/Df04 温馨/a/Ed33 和/c/Kc01 审美/n/Hj47 乐趣/n/Df12 。 /wp/-1

这项/r/-1 决定/n/Dk13 是/v/Ja01 [以色列/ns/Di02 农业部/n/Di09]nt 经过/p/Kb06 多次/m/Ka10 协商/v/Hi31 后/nd/Cb04 于/p/Kb01 日前/nt/Ca11 作出/v/Hc11 的/u/Kd01 。 /wp/-1

每个词的后面有两个标记，分别用斜线隔开。第一个标记是词性，第二个是词义标记。