2020年度媒体用字用语表说明

2020年度媒体用字用语调查语料来自2020年度国家语言资源监测语料库，包括平面、有声、网络三种媒体。平面媒体语料来源包括国内20家报纸2020年1月1日至12月31日的全部文本，共计**408,662**个文本，**419,584,239**汉字次。20家报纸按音序排列依次为：北京青年报、北京日报、北京晚报、长江日报、重庆日报、法制日报、光明日报、华西都市报、解放日报、今晚报、南方都市报、齐鲁晚报、钱江晚报、人民日报、山西日报、深圳特区报、新京报、新民晚报、羊城晚报、中国青年报；有声媒体语料来源包括中央电视台3个栏目、中央人民广播电台7个栏目以及央广网、北京、安徽、广东、山东、陕西、上海、四川等8家融媒体共37个栏目的转写文本，共计**66,266**个文本，**67,678,472**汉字次；网络媒体语料来自新浪的新闻网页，共计**400,088**个文本，**401,193,805**汉字次。三种媒体的语料量比例约为6:1:6。数据中，分别以A、B、C三个字母代表平面、有声、网络媒体。

**《2020年度媒体用字总表》**

2020年度媒体使用的汉字总字次为**888,448,937**，字种数为**10,970**，这些汉字与其在语料中的使用频次、文本分布、出现的媒体等属性一并形成《2020年度媒体用字总表》。其中覆盖整个语料80%的高频汉字共计557个，占全部用字种数的5.08%；覆盖整个语料90%的高频汉字共计877个，占全部用字种数的7.99%；覆盖整个语料99%的高频汉字共计2,247个，占全部用字种数的20.48%。

用字总表中频次低于10（包含10次）的汉字共计3,672个，仅使用1次的汉字共计1,346个。这些低频汉字多是罕用字，其出现情况大致可以分为以下几种：古书中的字句引用，或古文物介绍时用字；人名、网名中用字；方言字等。

与《2019年度媒体用字总表》相比，《2020年度媒体用字总表》中有9,581个汉字出现在2019年用字总表中，其余1,389个是2020年度的独用汉字。

（北京语言大学王莹莹、杨尔弘完成报纸、广播电视、网络新闻媒体的汉字、词语、成语等数据的统计、汇总；杨尔弘、腾永林、何婷婷分别提供了报纸、广播电视、网络新闻媒体的用字、用语基础统计数据）