
商务汉语口语话语标记标注规范

1 话语标记定义

口语交际中，独立于所在上下文句法结构之外的、不具有概念和逻辑意义、只具有话语交际功能的词、短语、小句，用于说话人对言语交际进行调节和监控。

注：此定义并不作为对于“什么成分才是话语标记”的硬性规定，主要是将话语标记与其他概念进行区分，方便更好地界定话语标记的特征。

2 话语标记判定依据

语料中的话语标记必须同时满足以下四个方面的全部要求。

（一）语用上

1. 只用于口语交际和以口语体或仿拟口语交际写成的书面语。
2. 如果去掉不影响语言陈述内容的概念、命题信息的完整性，同时也不影响其前后语境的逻辑和语义连贯关系。
3. 不能单独构成一个话轮，即不能独立用于交际。

（二）语义上

只具有程序性意义，不具有真实、明确的概念—命题意义。

（三）句法上：

具有独立性，不与相邻的句法成分构成更大的句法单位，即与相邻语言单位之间没有不存在组合或修饰关系，即使删除也不影响其所处句法结构的完整性与合法性。如“我说你能安静一会儿吗”中的“我说”与“你能安静一会儿吗”并不构成真正的句法结构关系。

（四）语音上

一方面倾向于弱化，常出现轻声、词尾语音延长以及音系缩减，前后常有停顿，但也可以没有停顿。

另一方面，话语标记是在“话语”层次上起作用的，因此对话语标记的分析也应在超句的话语层面上，故而话语标记需要在韵律上独立，后面允许出现停顿。

3 话语标记的标注角度

话语标记的标注包括**来源、语法类别、位置、韵律、功能**这五个维度。

（一）来源

1.1 标注角度

- （1）**专用话语标记：**专门用作话语标记的固定单位；

(2) **借用话语标记**：由已有词汇、短语或句子“虚化”而来，当其作为现实的语言单位来使用时具有字面上的实在意义。

例如“说白了”，在字面上无法解释，只能在语用上引入对前序话题的解释或评价，视为专用话语标记；“说实话”，字面上有“如实陈述”的意思，有作为非话语标记使用的相应语境，如“你给我说实话”。

1.2 标注方法

分类	概念	符号	例
专有	专门用作话语标记的固定单位	ZY	说白了
借用（兼类）	由已有词汇、短语或句子“虚化”而来	JY	所以、然后、你知道吗

(二) 语法类别与位置

2.1 标注角度

语法类别：词汇型、短语型、小句型

位置分布特征：定位（前置、中置、后置）、**不定位**

注：定位一类根据话语标记所依附和指向的话语序列，前置话语标记固定位于其依附和指向的话语序列之前，中置话语标记固定位于其依附和指向的话语序列之间，后置话语标记固定位于其依附和指向的话语序列之后。

2.2 标注符号

一级分类	二级分类	三级分类	符号	例
结构形式	词	语气词	WM	嗯
		代词	WP	那个、这个
		连词	WC	然后
		形容词	WA	好
		动词	WV	（你）看
		名词	WN	
	短语		PH	不是
	小句		ST	你知道吗
线性分布	定位	前置	Q	DM+S1+S2……
		中置	Z	S1+DM+S2
		后置	H	S1+(S2+S3……)+DM
	不定位		F	在标注中不使用

（三）韵律

3.1 标注角度

（1）停顿：与前后语言单位的时间间隔大大长于同小句中其他语言单位之间的间隔。发生在话语标记词汇上的停顿以符号“T”表示。发生在非话语标记词汇上的停顿用符号“(－)”表示。

（2）叠连：同一个话语标记在同一个语境下的叠用。

（3）延音：话语标记尾音的延长，如“那个...”。

（4）重读：为了突出最重要的内容或者叫信息中心，从而把相应的音节读得长些、高些，若要进一步突出，就再读得重些。重读音的声学特征是：调形较完整，音长比非重读音节长一半还要多，音高也较非重读音节高，或者说调域也较非重读音节宽。

（5）弱读：在口语中，有些音的发音可能变得比原来更弱更轻，这种现象即弱读。如轻声，很多就是典型的声调弱化，

3.2 标注符号

分类	符号	实例
停顿	T	那个就是，那个，那、就那些东西都有，我们家就没有。
叠连	L	我就一直认为我们家那个、那边就没有这个这个功能……
延音	Y	就是那些点播节目，还有什么视频影视之类的，那些是需要付费吗？
重读	Z	
弱读	R	

（四）功能：

4.1 标注角度：

（1）话语组织功能：组织话语、构建交际语境、保持话语意义连贯的能力。

（2）元语言功能：对自己的话语进行反馈、阐述、评价和修正。

（3）人际互动功能：唤起听话人参与、调节双方关系和态度等。

注：话语标记的功能分类具有层级性，本研究具体分类系统见 4.2。

4.2 标注符号

一级分类	二级分类	三级分类	概念	符号	例
话语组	交际进	交际启	提示一次言语交际行为的开始	ZZJQ	（节目即将开始，观众掌声将结束） 主持人：好，下面我们正式开始今天的……

织 功 能	程	动			
		交 际 转 换	用于改变交际对象、话题等	ZZJZ	客服：有线电视正常。 好 ，女士，您说一下，您的机顶盒盒号跟智能卡的卡号，您方便吗？ 客户：噢，是在机顶盒上面贴的那个吗？
		交 际 结 束	提示一次言语交际行为的结束一般后接道谢、道别	ZZJJ	客服：好的。 客户：嗯，好好，谢谢， 好 ，再见。
	话 轮 控 制	话 轮 转 接	用于原听话人承接或主动获取话轮	ZZHZ	客户：留的 1358169xxxx，就是我打的这电话。 客服： 那 地址给您对一下，是西城北三环中路 xx 号楼 x 门 xxx。
		话 轮 延 续	表示话没说完，还要继续	ZZHY	客服：嗯...那些手续您都知道了吗？ 客户：知道啦，带那个...就办理停机那个...那单子， 然后 带上户口本儿、身份证。
		话 轮 补 充	为说话人组织语言、寻找合适表达提供时间	ZZHB	客户： 嗯 ，你等会儿，我看一下啊。我可以，我可以选择。 嗯 里边哈我再找一下。视频 1 可以，高清电视可不对， 嗯 ..
元 语 言 功 能	自我反 馈	对自己说的话进行确认	YYZF	客户：塞一下， 对 ，要是能看，恢复了，我今儿好能看电视，要不然，看不了电视啊，那个闹的啊，一打开就是四个字母，什么 B00，B 呀还是 E 呀，我也不知道。	
	信息凸 显	对将要说的可能超出听话人预期的信息进行提示，强调新旧信息的对比	YYXT	客服：噢，那有可能会影响机顶盒了， 就是说 ，这边、我查一下也给您联系给这个维修部门那边了，然后他这边肯定会先给您查一下线路方面。 客户：噢，他明天什么时候过来呢？	
	解释说明	通过转换表达方式、解释具体内涵、举例等方式进一步解释刚刚说过的话	YYJS	客服：那您看，之前您联系我们这儿报修过吗？ 客户：呃，报修过。嗯... 是这样的 ，就是我们家这是，原来那个，就是，歌华搞活动抽奖得的机顶盒，长虹的那个。	

	表明态度	用于表达态度和情感	YYBT	客服：对，如果实在不行，只能联系这个维修给您查查了。 客户：噢，不是，他是正常看，就换台的时候干扰大，不换台、还换完台干扰的声音就有那交流声儿。
人际互动功能	联络关系	①言者唤起听者的关注或回应 ②听者的反馈。	HDLL	①客户：您能告诉我哪八家吗？抱歉，我问你。 客服：嗯，北京、建设，嗯，中国邮政、农业，华夏、工商，北京农村商业银行，现在加了一个交通银行…… ②客服：那先生，您能说下您机顶盒的盒号吗？ 客户：可以。稍等。盒号：0309，0V（客服：嗯。）1011，0V（客服：嗯。）1202，0V（客服：嗯。）9146。
	寻求认同	拉近与听话人的关系，使说话人的信息更容易被接受	HDQT	客户：我，我现在在试。我刚才都看得老好的，就是我这个重启，啊，不是，我中间重启了一下它就不行了，你知道吗。我看一下啊。
	退步协商	暗示将向说话人提供建议，强调自己并非要把建议强加给对方	HDXS	客服：您看要说这个的话，建议您可以说动动线，您看一下哪儿，连得对不对，您看可以做简单的连线，如果还是提示没信号的话，您这个机顶盒……

4 语料标注基本信息与格式（Office Word 格式）

任一语篇的标注文本都包含四个方面的内容：1. 语料基本信息；2. 特定类型语料信息；3. 谈话者信息；4. 标注工作信息；5. 文本内容

1. 语篇基本标注格式

在 Micro Word 环境下，任一语篇的基本标注内容与格式如下（“//”后为注释）：

[Text-ID]	//ID 是指该语篇的序号
[TextInfo-Style\CorpSort\Scene]	//语料基本信息，带有语体、语料类型和会话场景三个属性。
[OriFile]***[/OriFile]	//源文件，如原始音频文件
[Source]***[/Source]	//来源，即来自何处（网址、数据库或个人）
[/TextInfo]	
[SpecCorp-Domain]	//特定类型语料信息，带 Domain 属性
[Purpose] *** [/Purpose]	//交际目的，如维修、安装、费用
[/SpecCorp]	

[SpeakersInfo-Number]	//语篇参与人信息
[Speaker-IDN\ID\Sex\Age/]	
	//IDN 为说话者序列号，与会话者信息库中的会话人序列号相对应,这里可以填写为“音频文件名+S 或 C”；ID 为说话人，取值为“S、C、A”等。当描述脱口秀等单一会话人语篇时，ID 取值为“A”。注意结尾用“/”结束。
[/SpeakersInfo]	
[TagInfo]	//标注工作信息
[TagsBY] *** [/TagsBY]	//标注者。汉字全名，如：张三
[DateBY] *** [/DateBY]	//标注日期，格式为:YYYY-MM-DD
[Reviser] *** [/Reviser]	//校对者
[DateRev] ***[/DateRev]	//校对日期。
[/TagInfo]	
[Content]	//会话文本内容
[S]	//会话者 S
[Sent]	//Sent 表示句子，内含 DM 元素。Word 格式下无须标注。
[CH-Speaker]	//表示插话，Speaker 表示插话者，可以用 S 或 C 表示。插话中也可以包含话语标记 DM。
[DM-Orig\Cate\Posi\Rhyt\Func]***[/DM]	
[/CH]	
[DM-Orig\Cate\Posi\Rhyt\Func]***[/DM]	//DM 表示话语标记，带有属性 Orig、Cate、Posi、Rhyt、Func，分别表示来源、语法类别、位置、韵律、功能。如标注存有疑问，在 Func 后用英文“:”隔开，然后写可能的类型,或只写“CY”。如 ZZHZ:ZZJJ, 或 ZZHZ:CY
[/Sent]	
[/S]	
[C]	//会话者 C
[Sent]	//句子标记，不用人工标注，后期程序自动添加。
[DM]***[/DM]	
[/Sent]	
[/C]	

[A] //语篇作者 A, 或第三位谈话者
[Sent]
[DM]***[DM]
[/Sent]
[/A]
[/Content]
[OriText] //语料原文
[S]……[/S]
[C]……[/C]
[/OriText]
[/Text]

注释:

1) 语体: 分为口语、书面语, 分别以 0、1 表示。

0 口语

01 谈话语体

02 演讲语体 (含脱口秀等)

1 书面语

11 事务语体

要求简要平实, 是机关团体以及人们之间相互处理事务时运用的一种语体, 又分为公文体和日用体, 公文体包括行政公文、法律文书、财经文书等, 日用体包括书信、告白、条据等

12 科技语体

强调精确明晰, 是记载、传播社会科学和自然科学研究成果时运用的一种语体, 主要是各类科技学术作品, 包括自然科学和社会科学

13 政论语体

侧重逻辑严密, 是论述政治、社会问题, 进行政治鼓动时运用的一种语体, 主要是各种政治、评论、社论类的文章

14 文艺语体

以表现为主, 无论从选词造句, 还是设格增趣, 都须讲究文采, 主要指各类文学作品

15 新闻语体: 新闻时事报道

(若文本中包含有部分访谈、对话形式的内容, 则在相应语体后加数字标记 0)

2) 口语语料类型: 从三个角度标注: 对话 or 独白 or 记叙, 以 A1、A2、A3 表示; 公开 or 私人, 以 B1、B2 表示; 有准备 or 无准备 or 半准备, 以 C1、C2、C3 表示。如语料类型为对话、私人、无准备, 则标注为 A1B2C2。

- 3) 话语场景：分为 S1 自然会话、S2 电视访谈、S3 电话交谈、S4 医疗会话、S5 课堂会话、S6 市场促销、S7 商务会谈、S8 庭审会话、S9 其他
- 4) Domain 属性，即特定领域，分为 D1 电话客服、D2 商场销售、D3 商务谈判、D4 电视节目，等。
- 5) Speaker 元素带有四个属性，即 IDN、ID、Sex 和 Age 属性。IDN 属性指谈话者或语篇作者编号，用“语料序号+S”来标记；ID 的值为“S、C、A”等；Sex 属性表示性别，以 M、F 指代性别 Sex；Age 为年龄属性，分别以 Y、M、O 分别指代青少年、中年、老年。年龄属性值为估算值。
- 6) S、C、A 分别表示谈话者 S、谈话者 C、语篇作者或第三位谈话者 A。
- 7) DM 指话语标记 Discourse Mark，DM 后面加简单减号“-”，来与后面的标注符号分开。来源、语法类别、位置、韵律、功能均用上文规定的符号进行标注，之间用“\”分开。
- 8) 韵律特征包括停顿、叠连、拖音、重读、弱读，“N”表示无。如“你说那个那个——我到底要不要”的“那个”标为“LY”，即叠连和延音。另外，对于在句首出现的话语标记，其后不与话语单位相邻的话，才算有停顿；对于在句尾出现的话语标记，其前不与话语单位相邻的话，才算有停顿；对于在句中出现的话语标记，其后不与话语单位相邻，才算有停顿。
- 9) 对于存疑的话语标记，在功能标记最后标注英文“:”，然后写存疑内容，如 :ZZJJ，或 :CY。CY 表示存疑。

5 录音语料转写符号

对录音材料中的短暂停顿、延音现象、语音模糊等情况作相应记录，标记如下：

符号	意义	例
...	音节尾音的延长	嗯，那您看，要是说... 还是说...
(-)	小句中短暂的停顿	哦，您是您家那个 (-) 没信号是吗？
XXXX	转写者不能听清、确认的录音内容	对，你要不安那个，就不在进行状态中转，而且就跟以前使那 XXX，还得来回两边切换。
CH	插话	[S][Sent] 那个是四八分量线，[CH-C] 啊？[/CH] 一般您要插那跟线效果会更好一些。[/Sent][S]

六、语料转写标注示例（Office Word 环境下标注）

[Text-0000]
[TextInfo-01/A1B2C2/S3]
[OriFile]1439357.mp3[/OriFile]
[Source]歌华有线客服电话录音[/Source]
[/TextInfo]
[SpecCorp-D1]
[Purpose]咨询安装[/Purpose]
[/SpecCorp]
[Speaker]
[Speaker-1439357S\S\F\Y/]
[Speaker-1439357C\C\F\M/]
[/Speaker]
[TagInfo]
[TagsBY]*** [/TagsBY]
[DateBY]2022-09-30[/DateBY]
[Reviser]***[/Reviser]
[DateRev]2022-10-21[/DateRev]
[/TagInfo]
[Content]
[S]您好，歌华有线。[/S]
[C]哎，你好。[/C]
[S]嗯。[/S]
[C]呃，我们是在那个丰台新发地这儿。[/C]
[S]嗯。[/S]
[C]想安一个有线电视。[/C]
[S]您那小区叫什么名字呀？[/S]
[C]我们这儿不是小区，是，天伦锦城，挨着天伦锦城呢。[/C]
[S]那您这儿挨着，您这儿、您那儿叫什么呀？[/S]
[C]呃，啊，北京天然气。[/C]
[S]北京天然气？[/S]
[C]对。[/C]
[S]北、北京天然气什么呀？[/S]

[C]南郊输配站。[/C]

[S]您这是公司呀？还是，是单位吗？[/S]

[C]对，是单位的。[DM-JY\WA\F\T\YYZF]对[/DM]，单位。[/C]

[S]单位，单位的话，您得联系一下营业厅问一下，看您这儿能不能装了。[/S]

[C]我，我们这儿，我们[DM-JY\WP\F\N\ZZHB]那个[/DM]有装的。[/C]

[S]不是，您不是说您这是单位吗？[/S]

[C]我们是在单位的...我们属于[DM-JY\WP\F\N\ZZHB]那个[/DM]单位外边的家属院。[/C]

[S]您这家属院叫什么呀？北京天然气厂宿舍？还是什么呀？[/S]

[C]哎，对对对。[/C]

[S]就叫天然气厂宿舍？[/S]

[C]对。[DM-JY\WC\Q\N\ZZHZ]就是[/DM]咱们这边儿附近的营业厅哪有啊？[/C]

[S]您这儿是属于哪个街道或者哪个派出所啊？[/S]

[C]我们这边[DM-JY\WC\F\N\YYXT]就[/DM] [DM-JY\WP\F\N\ZZHB]那个[/DM]新发地批发市场这一块儿，这、这属于哪个派出所啊？[/C]

[S]我这边是通过这个街道或者派出所的名称划分的营业厅，您那儿属于哪儿啊？[/S]

[C]义前营啊？[DM-JY\WM\F\T\YYZF]啊[/DM]，义前营派出所。[/C]

[S]义前营？[/S]

[C]噢，是义前营。[/C]

[S]我这儿没有查到这个信息，这样吧，我给您念一下，这丰台区有好几个营业厅，您看您离哪个近吧。[/S]

[C]噢，您说。[/C]

[S]嗯，这儿有科技园、卢沟桥、马家堡、迎港和方庄。[/S]

[C]嗯，马家堡。[/C]

[S]离着马家堡近，是吗？[/S]

[C]对。[/C]

[S][DM-JY\WC\F\N\YYXT]就是[/DM]您确定您离这、离这近，那我就给您那个马家堡营业厅电话，您打电话问一下吧。[/S]

[C]马家堡，好，马家堡的，马家堡这边儿。马家堡离得近啊？花乡没有吗？花乡有吗？[/C]

[S]只有刚才我念的这几个营业厅，您那儿是不是属于新村儿啊？

[/S]

[C]什么？[/C]

[S]新村街道办事处啊？[/S]

[C]啊，我们这属于花乡。[/C]

[S]那您看看刚才这几个营业厅，您说您就离马家附近是吗？[/S]

[C]啊。[/C]

[S]那我就给您马家堡营业厅电话，您记一下。[/S]

[C]噢，您说。[/C]

[S]嗯，87592055。[/S]

[C]您慢点，875？[/C]

[S]87592055。[/S]

[C]2055，[DM-JY\WA\F\T\ZZJJ]好[/DM]。[/C]

[S][DM-JY\WA\F\T\YYZF]对[/DM]，您打电话问一下，看您的地址能不能装。[CH-C]哎，好。[/CH][DM-JY\WC\Z\N\ZZHY]然后[/DM]，这如果能装的话，我告诉您一下手续吧？[/S]

[C]哎，行，好的。[/C]

[S]嗯，拿着登记人的身份证，还有跟这身份证姓名相符的，指定银行的缴费存折或者银行卡。[/S]

[C]好。[/C]

[S]像北京、建设、中国邮政，这些大的银行都行。但是中国银行只能用存折，华夏银行只能用卡。您这儿楼房平房啊？[/S]

[C]工商银行行吗？[/C]

[S]嗯，工商银行的可以，您这儿是楼房平房啊？[/S]

[C]我们这儿平房。[/C]

[S]平房的话，您还拿得320块钱的初装费。[/S]

[C]噢，拿320的初装费？[/C]

[S]对。[/S]

[C][DM-JY\WP\F\N\ZZHB]那个[/DM]工商银行的存折行吗？[/C]

[S]存折和卡都可以。[/S]

[C]噢。[/C]

[S]然后您再打电话问一下，看您的地址是不是属于这营业厅，看这儿能不能装。[/S]

[C]好的。[/C]

[S]哎，好的。[/S]

[C]今天上班吗？[/C]

[S]今天上班。[/S]
[C]噢。[DM-JY\PH\Q\T\ZZJJ]好的[/DM]，谢谢。[/C]
[S]哎，不客气，感谢来电，再见。[/S]
[/Content]
[OriText]
.....
[S]新村街道办事处啊？[/S]
[C]啊，我们这属于花乡。[/C]
[S]那您看看刚才这几个营业厅，您说您就离马家附近是吗？[/S]
[C]啊。[/C]
[S]那我就给您马家堡营业厅电话，您记一下。[/S]
[C]噢，您说。[/C]
[S]嗯，87592055。[/S]
[C]您慢点，875？[/C]
[S]87592055。[/S]
[C]2055，好。[/C]
[S]对，您打电话问一下，看您的地址能不能装。[CH-C]哎，好。
[/CH]然后，这如果能装的话，我告诉您一下手续吧？[/S]
[C]哎，行，好的。[/C]
.....
[/OriText]
[/Text]
