

中医临床主诉中切诊描述的分词规范

望诊、闻诊、问诊和切诊（合称“四诊”）是中医专家在临床诊察疾病的基本方法，是在长期的医疗实践中，逐步形成和发展起来的。四诊分别从不同的角度诊察和收集患者的病情资料，各有独特的方法和依据^[1]。

中医临床主诉是对病人病情的高度提炼，是医疗行为的核心，是中医临床思维的起始和根基^[2]。中医临床主诉是中医专家通过四诊获取病人当次就诊最主要病情信息的记录载体。结构化中医临床主诉对医生更深入地调查、认识、分析、处理病人的病情具有重要意义。

在四诊中，切诊包括脉诊和按诊两部分，是在历代医家的医疗实践基础上，逐步补充和完善形成的极具中医特色的临床诊察疾病的方法，它在对病人的中医临床诊疗信息的收集起着重要的作用，其内涵丰富，需要耐心细致的体察^[1]。因此，围绕中医临床主诉中切诊描述的分词问题制定了本规范。

本规范根据切诊任务（即脉诊和按诊）将切诊描述的分词规则进行了分类，并参考了《中医四诊操作规范-第4部分：切诊》中的定义，借鉴了《信息处理用现代汉语分词规范》、《973 当代汉语文本语料库分词、词性标注加工规范》、《The Segmentation Guidelines for the Penn Chinese Treebank》、《中文临床记录分词规范》等规范，制定了对中医临床主诉中切诊描述的具体中文分词规则。

1、中医临床主诉中切诊描述中文分词总则：

（1）结合紧密原则：切诊描述中含有大量涉及医学符号、缩写及术语组合的描述，因此，在切诊描述分词规范中，使用“结合紧密”来处理这类词语是否需要分词。例如，在按诊中“周围硬块约 5*5cm”，其中“5*5”整体描述硬块组织大小，结合紧密，具有基数词（CD）含义，视为整体不作切分。本文的切分规则不仅考虑了词汇的基本意义，也考虑了数字、符号等医学元素的特殊含义。

（2）子词需有词性意义原则：切诊描述中含有大量中医脉象术语、病症名及人体部位名等，细粒度切分需要考虑切分后的子词是否含有词性意义。例如，在切诊中“脉律”意为“脉搏的节律”，加入助词“的”意思不变，故切分为“脉 律”。这些术语在中医临床诊断中具有重要意义，切分后各个子词均有词性意义，因此在切分时需严格遵守领域知识的约束条件。

1、脉诊描述的分词规则：

（1）方位词与“脉”的结合：

1) 单字方位词与“脉”结合时应切分。

例如：

- 在“左脉”中，“左”是方位词，因此，“左脉”应分词为“左 脉”。

2) 方位词后接“侧”，“边”，“面”等后缀时，整体不切分。

例如：

- “脉右侧略浮”，其中“右侧”视为整体不切分。

（2）修饰“脉”的状态的形容词与“脉”之间需要切分：

例如：

- 在“脉细”中，“细”是修饰“脉”的状态的形容词，因此，“脉细”应分词为“脉 细”。

（3）两个或两个以上用于修饰“脉”的状态的形容词之间需要切分：

例如：

- 在“脉细缓”中，“细缓”是由“细”和“缓”两个用于修饰“脉”的状态的形容词组成，因此，“脉细缓”应分词为“脉 细 缓”。
- 在“脉弦细数”中，“弦细数”是由“弦”、“细”和“数”三个用于修饰“脉”的状态的形容词组成，因此，“脉弦细数”应分词为“脉 弦 细 数”。

（4）修饰“脉”的状态的形容词与表示其程度的副词之间需要切分；结合紧密、使用稳定的不作切分：

例如：

- 在“脉略数”中，“略”是用于修饰“脉”的状态的形容词“数”的程度的副词，因此，“脉略数”应分词为“脉 略 数”。
- 在“脉弦缓解”，“舌脉如上”中，“缓解”，“如上”是结合紧密、使用稳定的修饰词，故不作切分。

（5）关于“脉”的相关术语：

1) 若能在词之间加入助词“的”且意思不变，则作切分，否则不切分。

例如：

- 在“脉律”中，“脉律”意为“脉搏的节律”，加入助词“的”意思不变，

故作切分。

2) 若能在词之间加入“和”等连词时且意思不变，则作切分，否则不切分。

例如：

- 在“筋脉”中，“筋脉”意为“筋骨和脉络”，加入连词“和”意思不变，故切分成“筋 脉”。

3) 整体呈词性的“脉”术语不作切分。

例如：

- 在“寸脉”，“关脉”，“尺脉”，“急脉”中，视为整体术语，不作切分。

4) 脉的相关缩写术语，缩写若能还原且还原后的术语存在临床记录中则切分。

例如：

- 在“脉结代”中，其中“结代”是“结脉”和“代脉”的并称，故切分成“脉 结 代”。

2、按诊描述的分词规则：

(1) 部位与修饰部位的形容词之间应切分开：

1) 单字形容词应与部位切分开。

例如：

- 在“手冷额热”中，“冷”和“热”分别形容“手”和“额”，故应切分成“手 冷 额 热”。

2) 两个或两个以上用于修饰部位的形容词之间需要切分。

例如：

- 在“前列腺肥大”中，其中“肥”，“大”分别形容“前列腺”的状态故切分为“前列腺 肥 大”。

(2) 修饰部位状态的形容词与表示其程度的副词之间需要切分：

例如：

- 在“腹略痛”中，“略”是修饰“痛”的程度副词，故应切分成“腹 略 痛”。

(3) 形容词+“性”或形容词+“状”的组合不作切分：

例如：

- “多发性”、“条索状”等不切分。

（4）数字：

1) 描述肿瘤等组织大小时，数字与“*”结合在一起时不作切分。

例如：

- “周围硬块约 5*5cm”，其中“5*5”描述组织大小，故不作切分。

2) 专业术语字母简写与数字结合在一起时切分。

例如：

- “L3”，其中“L”是腰椎（Lumbar）的缩写，“3”代表腰椎第三节，故切分为“L 3”。
- “P52”，其中“P”是脉搏（Pulse）的缩写，故切分为“P 52”。

3) 计量单位字母和数字结合时，应切分。

例如：

- “5cm”，其中“cm”是计量单位（Centimeter）的缩写，故切分为“5 cm”。

（5）量词

表示事物的单位或动作的量。如动词量词（次，趟，回），度量词（克，公斤，厘米，度），不定量词（点，些），个体量词（付），时量词（天，小时，分钟）各类量词均要切出。

1) 动词量词。

例如：

- “48 次”，其中“次”是动词量词，因此切分为“48 次”。

2) 时量词。

例如：

- “每分钟”，其中“分钟”是时量词，因此切分为“每 分钟”。

（6）方位词与部位的结合：

1) 单字方位词与部位结合时，观察方位词是否能替换，若能替换且替换后的术语存在临床记录中，应切分。

例如：

- 在“右手”中，“右”是方位词，且能替换成“左”，“左手”存在于临床记录中，故“右手”应分词为“右 手”。
- 在“腋下”中，“下”是方位词，但“下”被替换后，术语不存在于临床

记录中，故“腋下”视为整体，不作切分。

2) 方位词后接“侧”，“边”，“面”等后缀时，整体不切分。

例如：

- “右侧胸椎”，其中“右侧”视为整体不切分。
- “右侧面颈肩肌肉发紧”，其中“右侧面”视为整体不切分。

3) 部位名称后接“区”等表示区域的单字时，作为一个切分单位

例如：

- “心前区”，意思为“胸骨左缘三四肋间或左侧乳头附近的区域”。故视为整体不作切分。

(7) 特定术语在切分时，首先判断完整含义，再根据子词是否含有分类意义进行切分。

1) 若子词切分会产生歧义或不具有明显的分类意义，术语整体呈词性使用时，不作切分。

例如：

- “足背”，其中“背”容易引发歧义并且“足背”术语整体呈词性故不作切分。

(8) 形如“动词+形容词”的组合，判断“形容词”是否可替换：

1) 形容词可替换时，切分。

例如：

- “发紧”，其中“发”是动词意为“表现”，“紧”是形容词可替换成“软”，“硬”等词且替换后的词存在临床记录中，故切分成“发 紧”。
- “压痛”，其中“压”是动作，体现为“按压”，可替换为“按”等词且替换后的词存在临床记录中，故切分为“按 痛”。

2) 形容词不可替换时，不切分。

例如：

- “升高”等不切分。

(9) 形如“数词+名词”的组合。

1) 序数词应单独切分开。

例如：

- 在“二型糖尿病”中，其中“二”是序数词表示“第二种”故切分为

“二 型 糖尿病”。

2) “基数词+名词”的组合中，若插入量词后不改变原意，则进行切分。

例如：

- 在“四肢”中，插入量词后改变原意，故视为整体不进行切分。
- 在“脉迟，每分钟 40 左右”中，其中“40 左右”加入量词“次”，不改变原意，故进行切分。

3) “限定词+名词”两部分全为单字词，并且其中一部分属于界限词，则不切分。

例如：

- 在“全身”，“半身”中，“全”，“半”属于界限词，故不作切分。

参考文献

[1] 邓铁涛. 中医诊断学（修订版）. 上海科学技术出版社，2013.

[2] 周小青，黄惠勇，刘旺华. 中医主诉诊疗学. 中国中医药出版社，2017.