



方法论 独立日 日语日语 新工具 材料编

《NHK 日语发音音调新辞典》

MARCH 30, 2017

时隔 18 年，日本放送协会 NHK 于 2016 年 5 月推出了最新版的《NHK日语发音音调辞典》（下文称《新辞典》）。



目前还只有纸质版可购，尚无电子版。

《新辞典》的主要更新包括：

- 收录词条数从约 6 万 9 千增加至约 7 万 5 千。
- 采用了不同以往的音调记号。
- 根据对约 500 名 NHK 播音员的问卷调查变更了约 3500 个条目的音调。
- 丰富了附录中的解说和资料。

具体可观看「《新辞典》要点解说会」录影。

本文以下内容主要是对《新辞典》附录中部分内容的编译，包括：

- 《新辞典》采用新音调记号的缘由
- 日语音调呈现出的一些规律
 - 「降调」在什么位置不容易发生
 - 各类实词（和、汉、外来）原形的音调规律
 - 实词（名、动、形容）尾部接上虚词成分乃至发生变形时的音调规律（待更新）

本文与以下几期播客节目联动：

- #52：续·日语语音精要 | NHK日语发音音调新辞典
- #53：续·日语语音精要 | 新辞典·续

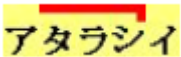
不同以往的音调记号

过往我们比较熟悉的音调记号大体有两种：

- 日本国语辞典《大辞林》及许多日语教材所采用的数字标记法：用数字表示在该单词的第几拍后面会发生降调。

◦ 《例》 ^{あた}「新しい」标 4 调，表示在第 4 拍「し」之后降调。

- 上一版《NHK日语发音音调辞典》所采用的上划线标记法：将单词完全用片假名写出，然后在音调处于高位的假名上方划线，并在降调处拐弯。

◦ 《例》 

《新辞典》所采用的音调记号其实与《大辞林》类似 ——

- 只不过它不是用数字，而是用记号「\」标明降调发生的位置：

◦ 《例》 アタラシ \ イ

- 若该单词中不发生降调，则在末尾加上记号「- -」：

◦ 《例》 アカイ - -

根据《新辞典》编者在附录中的解说，改变音调记号主要出于以下考虑：

- 「单词独立发音时，从第一拍到第二拍音调升高」是一个普遍现象，并不是单词本身的固有属性。

◦ 实际上，当单词前面接上修饰语时，「单词从第一拍到第二拍音调升高」这个规律可能被打破：

- 「男の子 | オトコノコ | 低高高低低」
- 「この男の子 | コノオトコノコ | 低高高高低低」

• 在实际自然语流中，每一拍的音调高低并不是二元的，即并不是只有“高”和“低”两种状态：

◦ 《例》 「変なズボンの男の子 | ヘ \ シナズボ \ シノオトコ \ ナコ」

- 整体音调一路走低：ヘ（高）→ シナズボ（低）→ シノオトコ（更低）→ ノコ（再低）

• 综上所述，对一个单词而言，重要的并不是「哪些拍高，哪些拍低」，而是「降调发生在什么位置」。

在下文中，我们会混用《新辞典》的记号标记法与《大辞林》的数字标记法。

另外，下述针对调型的称呼也较为常见：

称呼	数字标记法	《新辞典》标记法
平板型	0 调	アカイ — —
头高型	1 调	コ \ クド
尾高型	n 调 ⁰	オトコ \
中高型	其他	ブル \ —

「降调」不容易发生的位置

↑ 多与特殊拍相关

a. 紧跟着长音「ー」与拨音「ン」的位置，原则上不发生降调。

《例》

- 「カレ \ ーパン」，不是「カレ~~ー~~パン」
- 「メロ \ ンパン」，不是「メロ~~ン~~パン」

《例外》

- 「コー \ ン茶」

b. 紧跟着双元音 [-ai] 中的副音「i」的位置，倾向于不发生降调。

《例》

- 「総菜パン | ソーザ \ イパン」，不大说「ソーザ~~イ~~パン」
- 「ネクタ \ イピン」，不大说「ネクタ~~イ~~ピン」

c. 紧跟着无声化拍的位置，有时也不容易发生降调。

《例》

- 「共益費」的首选音调是「キョー工 \ キヒ」，次选音调才是「キョー工
キ \ ヒ」←「キ」被无声化

综上所述

降调（按照其他规律）本来应该发生的位置有时会由于某些音拍（长音、拨音、双元音副音、无声化拍）的存在而发生变动。在下文中，若有这样的情况发生，将在词例右上角标注星号*。

实词原形的音调规律

和语词

a. 形容词原形

降调基本都紧跟在倒数第二拍后面（即所谓 -2 调）。

《例》

- 「濃い | コ \ イ」
- 「うまい | ウマ \ イ」
- 「短い | ミジカ \ イ」

《例外》

- 在《新辞典》中以下形容词只标有 0 调（即平板型，不发生降调）：

- 「^{あつ}厚い（^{あつ}篤い）」
- 「^{おも}重い」
- 「^{あら}粗い（^{あら}荒い）」
- 「^{あか}赤い」
- 「^{かる}軽い」
- 「^{くら}暗い」
- 「^{とお}遠い」
- 「^{き いろ}黄色い」
- 「^{うす}薄い」¹
- 「^{かた}固い（^{かた}硬い、^{かた}堅い）」¹
- 「^{かた}難い」
- 「^{きつ}きつい」¹
- 「^{まる}丸い」¹
- 「いけない」²

总的来说，原本就属于少数派的 0 调（平板型）形容词原形正缓慢地向 -2 调合流。

b. 动词原形

【复合动词】

降调基本都紧跟在倒数第二拍后面（即所谓 -2 调）。

《例》

- 「移り変わる | ウツリカワ \ ル」
- 「食べ飽きる | タベアキ \ ル」
- 「学び取る | マナビト \ ル」

【2 拍动词】

0 调与 -2 调各半。

【3 拍以上的动词】

以 -2 调居多。

《例》

- 「味わう | アジワ \ ウ」
- 「炒める | イタメ \ ル」
- 「食べる | タベ \ ル」
- 「通る* | ト \ ール」
- 「入る* | ハ \ イル」

《例外》

- 「いただく | イタダク - -」
- 「固める | カタメル - -」

汉语词

a. 二字汉语（1 拍 + 1 拍）

多为 1 调（即头高型）。

《例》

- 「趣味 | シュ \ ミ」
- 「図書 | ト \ ショ」
- 「馬車 | バ \ シャ」

b. 二字汉语（1 或 2 拍 + 2 拍）

多为 0 调（即平板型）。

《例》

- 「希望 | キボ- -」
- 「珠算 | シュザン- -」
- 「愛犬 | アイケン- -」
- 「学習 | ガクシュ- -」

《例外》

- 「宇宙 | ウ \ チュー」
- 「理解 | リ \ カイ」
- 「支配 | シ \ ハイ」

c. 二字汉语（2 拍 + 1 拍）

表名词性含义时，多为 1 调。

《例》

- 「国土 | コ \ クド」
- 「博士 | ハ \ クシ」
- 「色素* | シキ \ ソ」³

表动词性含义时，0 调的也不少。

《例》

- 「飲酒 | インシュ- -」
- 「開始 | カイシ- -」
- 「食事 | ショクジ- -」

外来语

a. 三拍以下

多为 1 调。

《例》

- 「ガ \ ム」
- 「カ \ メラ」
- 「ポ \ スト」

《例外》

- 「カレー - - 」
- 「コップ - - 」
- 「ブル \ ー」

↑ 例外中以最后一拍为特殊拍（长音、拨音）者居多。

b. 四拍

约半数为 1 调，其余呈现以下规律：

- 第三拍为特殊拍时，多为 -3 调，《例》「ドリ \ ップ」「オリ \ ーブ」「ジレ \ ンマ」。
- 整体完全不含特殊拍时，多为 0 调，《例》「アルバム - - 」「オカルト - - 」「オムレツ - - 」。

c. 五拍以上

多为 -3 调。

《例》

- 「アスファ \ ルト」
- 「ストラ \ イク」
- 「イミテ \ ーション*」

《例外》

最终拍为长音时，倾向于变成 -4 调：

- 「エコ \ ノミー」

- 「カテ \ ゴリー」
- 「テクノ \ ロジー」

缩略语

名词 + 名词（3 或 4 拍）

多为 0 调。

《例》

- 「うな^{どん}井 | ウナドン [—] [—]」 ← うなぎ + どんぶり
- 「国連 | コクレ^ん [—] [—]」 ← 国際連合
- 「ラノベ [—] [—]」 ← ライトノベル

《例外》

- 「春^{しゅんとう} 闘 | シュ \ ントー」 ← 春季闘争
- 「夏^{なついち} 一 | ナツ \ イチ」 ← 夏の一冊

当词语作为缩略语的身份在使用者的意识中变得淡薄时，可能产生其他调型：

- 「通販^{つうはん} | ツ \ ーハン」 ← 通信販売

由单个单词缩略而成的词语不适用于本规则：

- 「テ \ レビ」 ← テレビジョン
- 「サ \ ンド」 ← サンドイッチ

实词接上虚词后的音调规律

（待续）

0: 在这里 n 等于该单词的拍数（如对「男^{おとこ}」而言， n 等于 3），表示降调在最后一拍之后发生（故又可称 -1 调）。-1 调单词单独发音时的音调与 0 调无区别，降调只在该单词后续接上虚词成分时才体现出来。↩

1: 列表中这 4 项在《スーパ-大辞林3.0》中标有 0 调和 2 调两种调型。↩

2: 「いけない」在《大辞林》等辞书中被归类为「連語」而非「形容詞」。↩

3: 此处「シキソ」中的「シ」为无声化拍，详见「降调」不容易发生的位置。
↩

日语电子辞典及日文输入法

OCTOBER 21, 2015

初学阶段，你可能只能无奈地使用收词量较少且释义难以切中要害的**日中辞典**。

- 小学館中日・日中辞典：iOS，安卓
- ↑ 该日中辞典已于 2015 年 11 月时隔 14 年刊行新版，上述安卓收费应用的内容暂未**相应更新**。
- 如果不愿意付费，可使用沪江小D，其优势是标出了单词的调型并具有发音功能，缺点是调型**可能有误**，发音可能是合成语音，且带有广告。
- 如果访问国外网站速度不错，亦可使用 weblio 提供的在线**日中・中日辞典**。

如果有一定的英语基础，可将**和英辞典**作为补充。

- 三省堂 WISDOM 英和・和英辞典：iOS，安卓
- ↑ 该辞典在 Apple 的 OS X 及 iOS 系统中有**免费附赠**。在 OS X「系统偏好设置」中的「语言与地区」中加入「日本語」即可在系统自带的「词典」应用中查询。在 iOS 中可通过「设置-通用-词典」添加，之后就可以通过「选中-查询」来查词了；如果想像普通辞典 App 那样使用，可以从 App Store 下载 **极光词典**。

- 免费且支持手写的Japanese App。
- 亦可使用 webllo 提供的在线英和・和英辞典。

但是你真的，真的，真的应该尽快过渡到使用**全日文辞书**的阶段。

- 三省堂超级大辞林：iOS，安卓
- ↑ 该辞典标出了单词的调型，且在 Apple 的 OS X 及 iOS 系统中有**免费附赠**，调用方法与上述和英辞典相同。
- 小学館大辞泉：iOS，安卓，Windows
- ↑ 小学館每年会对上述电子化版本的『大辞泉』做 3 次内容更新。
- 如果不愿意付费，可使用包含了『大辞林』、『大辞泉』等辞书的在线辞典 Kotobank 。

遗憾的是，上述辞典要么不具备**发音功能**，要么发音功能不够靠谱，这方面可选用——

- NHK 日语发音语调辞典：iOS，安卓
- ↑ 已于 2016 年 5 月刊行新版，电子版尚未更新。
- 三省堂新明解日语语调辞典：实体书 + CD
- 查询单词变形后的语调可使用在线日语语调辞典（OJAD）。

要查电子辞书，在无法复制粘贴的场合，你得先**输入日语**。

在桌面端，新近的 OS X 及 Windows 自带的日语输入法已足够好用。

- OS X：在「系统偏好设置」中的「键盘－输入源」中增加「日文」即可。
- Windows：在「控制面板－时间、语言和区域－语言」中增加「日本語」即可。
- ↑ 某些盗版 Windows 系统去掉了输入法部分，无法正常添加。
- Linux：可使用 Google 开源的 Mozc。

桌面端的日语输入法通常有平假名、片假名、罗马字这几种输入状态（在输入法状态条上分别以「あ」「ア」「A」来标示）。通常保持在平假名状态（「あ」）即可，有需要时，可以在输入平假名后——

- 按 F7 将其变换为相应的片假名。
- 按 F10 将其变换为相应的罗马字。
- 按 F6 可将其变回平假名。

- ↑ 具体快捷键可能略有出入，总之都是 F 键就对了。

在移动端——

- iOS：在「设置－通用－键盘－键盘」中添加「日文-罗马字」或/和「日文-假名」即可。
- ↑ 「日文－罗马字」的输入方式与汉语拼音是一个道理，但熟练后打字速度更快的输入法是「日文－假名」，其使用方法则可用 App「跟着接接学日语」学习。
- 安卓：可使用免费的 Google Japanese Input 或收费的 ATOK。

有时你会需要手写由于读音不明而不知如何输入的汉字。

- 上述移动端辞典 App 大多自带手写输入功能，若不具备——
 - iOS：可使用 MyScript Stack 或是用系统自带的「中文手写输入法」凑合。
 - 安卓：可使用 Google 手写输入。
- Windows 自带的日语输入法本身支持手写输入。
- 在 OS X 下则可选用免费的 Google Japanese Input 或收费的かわせみ 2。

有时你会弄不清句子该怎么拆分，导致辞书无从查起——

- 此时可借助自动分词工具（其一、其二、背景简介）

最后，你或许会从别处听闻被称作「日语辞典神器」的 EBWin，这里不作推荐的原因有：

- EBWin 读取的文件格式是 EPWING。在移动互联网时代之前，它曾是日本电子辞书的主流格式，但如今已逐渐式微。譬如「三省堂超级大辞林」最新的第三版就没有以 EPWING 格式发行过。
- EPWING 辞书的发售载体通常是实体光盘，购买正版相当麻烦。
- 又及，一个基于 EPWING 的有趣项目：青空文库全文 EPWING（背景简介）

少納言、青空 WING、活辞典和编舟记

OCTOBER 11, 2015

传统辞书（包括已经电子化了的）中的单词用例都是孤零零的句子。当你想在更大范围的现场观察这个词的使用时，会发现：没有语境。又比如你用日语写东西，想查下别人对某种搭配的使用。传统辞书以收词为主，即便收录搭配，限于篇幅也是挂一漏万，经常帮不上忙。一个很自然的想法是：用搜索引擎。然而，搜索引擎给出的结果往往良莠不齐且过于庞杂。更具可靠性和相关性的方法是：检索日语语料库。

日语语料库检索系统中较为常用的有「少納言」。它基于日本国立国语研究所（简称“国语研”）制作的“现代书面日语均衡语料库”。“均衡”讲的是它的收录面：涵盖书籍、报刊、博客、教科书、国会议事录等门类，时间跨度自 1971 年至 2008 年，总篇幅 1.43 亿词。作为「言の葉」计划的一部分，国语研还在构建一个“超大规模语料库”，篇幅预计为上述“均衡语料库”的 100 倍。除提供语境外，语料库还可以被用来辨析近义词。日本筑波大学开发的语料库检索工具 NLT 提供「2 語比較」功能，能从语料库中抽取 2 个词各自（相对）专属的搭配。

上述查询工具都基于网页，频繁使用时往往为网络延迟所累。这时你可以考虑——「青空 WING」。顾名思义，「青空 WING」是「青空文庫」与「EBWING」的结合。前者是日本最大的在线公有领域书库，后者是日语辞书在移动互联网时代之前最为常用的电子格式。

「青空 WING」的使用与普通「EBWING」辞书并无二致，但其内容却是「青空文庫」全文。这意味着你可以在本地随时对「青空文庫」收录的近 800 位作家的 12000 多部作品进行全文检索。「EBWING」系列词典软件（EBWin，EBMac，EBPocket）的开发者近来为软件增加了制作全文检索索引的功能，使得原本需要至少耗费数秒乃至十多秒的全文检索得以瞬间完成！

不过，上面讲到的「少納言」和「青空 WING」有一个通病：未对语料进行分词处理。这使得你搜「ひと」，会把「ひとつ」、「ひとまず」之类也给搜出来。

「少納言」的加强版「中納言」当无此问题，但其使用需专门申请。「青空 WING」的制作者另外提供的「作家別用例辞典」由于经自动分词程序 Mecab 预处理过，也在很大程度上规避了此问题。

今年 7 月 18 日,「青空 WING」的制作者发布了基于「青空 WING」的“青空文库用例卡片制作工具”。想来,这将大大方便电影《编舟记》所描写的,编撰辞书时的用例收集工作。与自动分词工具一样,语料库检索工具对语言学习者而言,也还远谈不上好用。但愿这只是一个开头。

分词、句子树、语言处理 100 击

OCTOBER 11, 2015

假如你正打算或刚开始学习日文,将会有有一个长期困扰你的问题。假如你学习日文已有一段时日,此问题或许仍时不时给你带来不快。它便是:分词。

英文中的词被空格隔开,几乎就不需要分词。中文里虽没有空格,作为最小表意单位的汉字却是天然分开的,查不了词典至少可以先查字典。而到了日文这儿,事情变得相当麻烦,譬如某日文版《小王子》开头讲“蟒蛇吞猎物”的一段。从中可以看出,日文同中文一样并不惯用空格。里面那些并非汉字的文字叫做假名。要命的是,日文中的最小表意单位往往是由**不确定数目**的假名连在一起构成的。想查词典吧,还得先搞清(或猜中)该从哪儿断开,用哪一串假名去查。猜错个几次,便坏了读书的兴致。

有没有办法自动分词呢?答案是:有。

人们试图用计算机为日语自动分词的尝试至少可以追溯至上世纪 1992 年京都大学对日语分词系统「JUMAN」的开发。1996 年,奈良先端科学技术学院(该学院的特点是强调前沿科研,只设研究生院,不设本科生院)在 JUMAN 的基础上开发了「茶筌(Chasen)」并将其开源。其后,JUMAN 和 Chasen 都不断迭代,并衍生出了如今最为优秀的日语分词系统「和布蕪(Mecab)」。

2010 年,日本国立国语研究所和千叶大学的研究者们从报纸、文学作品及博客中抽取了共计 3200 句日文,用 Mecab 配合他们制作的形态素解析辞书 UniDic 对这些句子进行解析,总体分词准确率超过了 99.4%。如今,在网络上不难找到基于 Mecab 的在线分词工具(其一、其二),粘入日语文本,获得分词后的结果,其中还包含对汉字读音和单词原形的推测。

然而，使用在线分词工具也好，在自己的电脑上安装配置 Mecab 也好，对只想安静自在地读点日文的学习者来说都远不够友好和便利。这自然是希望运用信息技术改善语言学习者体验的人士（包括笔者在内）所需要面对的课题。不过，除了使用现成的工具，你还可以走得更近些，看看在 Mecab 和 UniDic 的背后和周围正发生着什么。

分词任务从属于一个名为自然语言处理的交叉学科。苹果 iPhone 上的 Siri 等语音助理就是基于自然语言处理技术的产物。2010 年，上述奈良先端科学技术学院的乾健太郎副教授来到地处仙台的日本东北大学，创立了那里的自然语言处理研究室。2011 年，日本自然语言处理界的代表人物辻井润一教授从东京大学退休，任微软亚洲研究院首席研究员。311 震灾后，其门下的冈崎直观研究员赴仙台加入了乾教授的研究室，任副教授。

研究室的新人不一定经历过正统的编程训练。为了带他们入门，并使他们养成良好的编程习惯，冈崎老师编写并公开了一套名为语言处理 100 击的练习。重点是，这套循序渐进的练习全部选用极具实用性的题材！如：

- 第 31 题，对夏目漱石的小说《我是猫》进行自动形态素解析，从中抽出所有动词。
- 第 44 题，对《我是猫》中的内容进行自动句子结构解析，并用树状图将其展现出来。
- 第 70 - 79 题，用机器学习自动判断影评中的句子是在赞还是在踩。
- 第 80 - 99 题，以十余万篇维基百科词条为材料，用向量空间法让机器学习单词语义。

对已有轻度编程经验及少量日语基础的人而言，练“100 击”既能入门自然语言处理，又能加强日语及编程技能，一石二鸟。对不具备这些基础的同学而言，想从“100 击”获益，恐怕还得配上一套手把手的教程。这样的一套教程目前尚不存在，希望以后会有。

引用欢迎 | 转载随意