

詞型與詞形：你能分得清嗎？

——從詞彙語義角度看術語學應該規範什麼1

張化瑞

香港理工大學中文與雙語學系 香港
北京大學計算語言學教育部重點實驗室 北京
hrzhang@pku.edu.cn

摘要：在漢語中，同音異義詞的出現是無法避免的，但作為術語，特別是在一個非常狹窄的範圍內在某種意義上構成對立或互斥關係的術語就應該極力避免使用同音異義詞，特別是同音形似異義詞。“詞型”和“詞形”就是這樣的典型例子。

同音異義並舉詞是標準急需規範的，即進行分辨，給出不同的說法；同音異形同義詞如有理據則完全沒有規範的必要，或者說規範可以認定它們為同義詞。前者以“詞型”和“詞形”為例，後者以“固步自封”和“故步自封”為例。

結論為，語音形式與詞彙意義的一一對應關係的破缺，是術語標準最需要彌補的。

關鍵詞：同音異義並舉詞 詞型 詞形 語音形式 詞彙意義 術語規範化

'Word Type' and 'Word Form' in Chinese: Should Opposite Homophones Be Normalized in Terminology?

HuaRui Zhang

Department of Chinese & Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong
Key Laboratory of Computational Linguistics (Peking University), Ministry of Education, China
hrzhang@pku.edu.cn

Abstract: In Chinese, the emergence of homophones is inevitable, however, opposite homophones should be avoided with best efforts, especially in a very narrow field.

The Chinese equivalent of "word type" and "word form" given by the PRC national standard GB/T17532-2005, which forms opposite homophones, is one of such typical examples.

Homophones need to be normalized most urgently, namely, being distinguished and given different phonetic forms; while homophone synonyms do not need to be normalized arbitrarily.

The conclusion is that the breaking of one-to-one correspondence between phonetic form and lexical meaning of opposite homophones needs to be normalized first in terminology.

Key words: opposite homophones, word type, word form, phonetic form, lexical meaning, terminology normalization

1 引言

1 本研究得到國家自然科學基金-重大研究計畫 項目號:91024009 的部分資助。

從概括到具體，“詞”的概念感覺應該分成三層（從量的角度進行區分）：

lemma: 如果集合是整個詞彙，則 good, better, best 只有一個有資格作為元素（通常選 good），元素個數為 1；

word type: 如果集合是整個詞表，則 good, better, best 每個都有且只有一個名額，元素個數 3；

word token: 如果集合是整個語料庫，則 good, better, best 每個都可能不止一個名額，元素個數為 $n_{\text{good}}+n_{\text{better}}+n_{\text{best}}$ (n 指其頻次)。

但我們看一看 GB/T17532-2005 《術語工作 計算機應用 詞彙》[1]中是怎麼定義的：

（讀該國家標準能明顯感覺到閱讀過程中自然語言帶來的不確定性。）

3.8 詞型 type

<語言工程>文本(3.6)中代表一個確定類別的語言單位。

獻疑：沒有舉例，“確定類別”不知道是如何確定的。

3.9 詞例 token

<自然語言處理>詞型(3.8)在文本中的具體表現形式。

獻疑：是一次出現算一個，還是一種形式算一個？

3.11 詞形 word form

給定詞的任何形態句法變體。

例：在英語中，indicate, indicates; visitor, visitor's.

注：在屈折語中，詞形經常是屈折形式，例如，英語的 go, goes; 以及 go, went, gone。

3.12 詞形變化範型 paradigm

<自然語言處理>屬於某一給定詞或詞組型術語的各個詞形(3.11)的類聚。

獻疑：與詞型(3.8)是什麼關係？與其構成等價關係？還是代表性元素與集合的關係？

3.13 基本詞形 base form

參照詞形 reference form

根據詞典編纂學的慣例選出的用以表示一個詞形變化範型(3.12)中的各個形式的詞形(3.11)。

例：在英語中，bind 是 bind, bound, binds, binding 等的基本詞形。

獻疑：看來 基本詞形(3.13) 是 詞形變化範型(3.12) 的表示方式。那 基本詞形(3.13) 與 詞型(3.8) 是什麼關係？

2 同異之間 問題凸顯

(一) 求同辨異

看了“詞型(3.8), 詞形變化範型(3.12), 基本詞形(3.13)”的定義, 感覺三者相互之間關係很密切, 但差別在哪裡不清楚。標準就應該是對概念/術語“求同辨異”的。說同, 又用了不同的術語, 且並未指明其為同義或別稱; 說異, 又未指明兩兩之間到底哪里有差別, 只能靠讀者去猜測或歸納: 自然就會出現不同的解讀和結果, 而這本來正是標準要規範的。

(二) 晦同顯異

自然語言發展出同音異義詞也就罷了, 術語特別是在一個非常狹窄的範圍內在某種意義上構成對立或互斥關係的術語就應該極力避免使用同音異義詞, 特別是同音形似異義詞。“詞型”和“詞形”就是這樣的典型例子。“期中”和“期終”是日常生活語言中的例子, 人們尚且知道用“期末”去代替“期終”以免說不清楚。“權力”和“權利”的說不清楚則反應的是造詞者當時根本就沒有這方面的意識、用詞者現在也沒有太迫切的需求去區分, 下意識地認為二者相差不多(有“權力”者享有更多的“權利”), 才會形成這樣的語言使用狀況。

歧義同音, 一定是人們覺得沒關係, 不影響表達, 比如數學中的“負數”和“復數”, 前者和“正數”對舉, 后者和“實數”對舉, 不在一個層次上, 根本無法形成對立或互斥關係, 所以不會影響表達的效率。

(三) 大同小異

“形”與“型”, 在意義上是有細微差別的, 但很多時候關係太密切, 不易劃清界限, 比如“臉型”和“臉形”, “流線型”和“流線形”。常見的還有“五筆字型”, 很多人誤為“五筆字形”, 因為有“形碼”之說。“型”抽

象一些，“形”具體一些。所以“五筆字型”的“字型”是指漢字的結構類型。但問題是“五筆字形”並不構成和“五筆字型”的對立，只不過是一個表達不當，一般人都能識別其真實的意思表示。“臉型”和“臉形”也是一樣，字形差別也不大，意思上差別也不大，甚至通用無妨。“流線型”和“流線形”在意義和使用上更是沒有差別，要不要把“流線型”定為區別詞，把“流線形”定為名詞？或者把其中一個定為別字或異形詞？

下面是 Google 搜索的結果數(網頁數，K:10³; M: 10⁶):

“流線型外形”	7.19K
“流線形外型”	1.32K
“流線形外形”	293.
“流線型外型”	4.3K
“流線型”	1.07M
“流線形”	207K
“外型”	7.69M
“外形”	24.9M
“臉型”	3.36M
“臉形”	280K
“體型”	15.8M
“體形”	3.66M
“字型”	6.72M
“字形”	5.08M (“字型” 和 “字形” 次數相近，是因為二者意義不同)

很多相差幾倍，最多在 10 倍這個量級，作為語言差異，不算懸殊。若以“例不十不立”的標準來說，則大大富餘了。馮志偉老師提出應使用概率的觀點[2]，我想至少也應該說“比不十不獨立”吧。就是說，在用法都很多的情況下，如果一種用法沒有超過另一種用法 10 倍這個數量級，不能認為其已占有壓倒性優勢，因而弱勢用法應該被認為是有問題的或錯誤進而加以禁止。

3 同音異形同義詞

由於字義的模糊性和交叉性，產生異形詞是非常自然的事情，有的在表達上明顯有優劣之分（不是指使用頻度），可以淘汰一個，但不少異形詞其實是兩個都講得通的，比如“固步自封”和“故步自封”，細分起來，也許可以說一個是空間上的停滯，一個是時間上的停滯。但實際使用中完全可以作為同義

詞。音不同可以同義，音相同就不許同義了嗎？多幾個同音異形同義詞，只要都有理據性，完全不會增加使用者的負擔（語言的複雜度），因為它是有理的；相反，一定要從都有理據的兩個中規範一個，禁止一個，則是額外的負擔，是非常容易記錯/用錯的，高考出題的人可能覺得有題可出了。描述相同的事物機理，規則並非越多越好，是越少越好。當我們的規則越來越多的時候，我們誤以為記住的規則(信息)越多知識越豐富，實際上是我們對事物的本質認識（也許這才可以稱為知識）在越來越少。

該國家標準中把信息定義為“知識”，也是值得商榷的。目前經常有“信息、知識、智慧”的對比說法，同時也有並非面向某種應用的通用定義[3]。

3.1 信息 information

〈信息處理〉關於客體(如事實、概念、事件、思想、過程等)的知識，它在一定的上下文中具有特定的意義。

注1：本條引自 ISO/IEC 2382-1。

注2：在其他的應用領域，信息的定義不同。

如果從這個角度來看詞類的問題，詞類分得越多，表面上看，是我們的知識越多，實際上不見得。如果反映的不是本質差異的話，徒增代價。

4 同音異義並舉詞

同音異義並舉詞是標準急需規範的，即進行分辨，給出不同的說法；同音異形同義詞如有理據則完全不用硬性規範，或者說規範可以認定它們為同義詞。

前者以“詞型”和“詞形”為例，后者例如“固步自封”和“故步自封”。

但二者之間的界限並不明晰，比如：

“訂金”和“定金”，很多人(包括現漢)認為二者是異形詞，其實二者的使用目前有相當大的區別，特別是法律上也有明文支持（“定金”買方違約不返還，賣方違約雙倍返還，有擔保金的意思）。

目前我國現行法律中對定金的規定主要有《民法通則》第 89 條第 3 款、《擔保法》[5]第 6 章、《合同法》第 115 條、116 條及《最高人民法院關於適用〈中華人民共和國擔保法〉若干問題的解釋》[6]第 6 部分。

定金擔保是有懲罰性的，《擔保法》第 89 條規定“當事人可以約定一方向對方給付定金作為債權的擔保。債務人履行債務后，定金應當抵作價款或者着收回。給付定金的一方不履行約定的債務的，無權要求返還定金，收受定金的一方不履行債務的，應當雙倍返還定金。”其中“無權要求返還定金”和“雙倍返還定金”的規定都凸顯了定金擔保的懲罰性。

而《最高人民法院關於適用〈中華人民共和國擔保法〉若干問題的解釋》第一百一十八條更明確規定：

當事人交付留置金、**擔保金**、**保證金**、訂約金、押金或者**訂金**等，但沒有約定定金性質的，當事人主張定金權利的，人民法院不予支持。

這說明了法律語言的嚴肅還是不嚴肅？法律語言的後果當然是嚴肅的，你只能接受（很多購房者被銷售者用此種手段騙去了數以萬計的說是“訂金”寫為“定金”的錢，“口說無憑，白紙黑字”）；但法律制定者使用這樣的“‘說’不清楚”的語言顯然是非常不嚴肅的，甚至應該說正是由於法律語言的不嚴肅造成了消費者的無辜受害（完全不是他們真實的意思表示）。因為在古漢語中並沒有“定金”和“訂金”的這一巨大區別，完全是近幾十年才發展起來、特別是由法律確定下來的（詞典中都未明確區分，而且認為它們是異形詞，也就難怪一般人意識中認為它們沒有多大區別）。本來就可以叫“擔保金”或其他意義明確、語音也不會混淆的名稱（但請注意，“擔保金”這一名稱已被法釋否定，將其認同為“訂金”），非常明確，沒有誤解的可能。因此，在這一點上說，沒有語音區別力的法律語言成了騙子們的護身符，是符合事實的，其責任是無可推卸的。

5 結論

上述同音異義並舉詞，可以說是語音形式與詞彙意義的一一對應關係的破缺。術語規範化的一個重要目的就是要彌補這種破缺。

參考文獻

- [1] 中華人民共和國國家標準 GB/T17532-2005《術語工作 計算機應用 詞彙》，2005 年
- [2] 馮志偉 《“例不十，法不立”原則和統計方法》，詞庫建設通訊，1996 年 8 月，總第 8 期，香港
- [3] 謝清俊，謝瀛春 《一個通用的資訊的定義》，信息科學交叉研究學術研討會，北京師範大學，2005 年 11 月 4-5 日

- [4] 中華人民共和國教育部、國家語言文字工作委員會 《第一批异形詞整理表》，2001 年 12 月 19 日
- [5] 中華人民共和國擔保法，新法規月刊，1995 年 08 期
- [6] 最高人民法院關於適用《中華人民共和國擔保法》若干問題的解釋，新法規月刊，2001 年 02 期