

分詞標準

一、定義：定義切分字串的基本單位是訂定分詞標準的首要工作

1. 詞：具有獨立意義，且扮演特殊語法功能的字串。
2. 基本詞類：動詞、名詞、副詞、定詞、量詞、介詞、方位詞、連接詞、語助詞、感嘆詞。
3. 細分因素：動詞、名詞、副詞、定詞、量詞難以判斷(動詞、名詞尤其)。
 - (1) 動詞和名詞另有詞組形式，因此擁有複合詞、詞組的問題，另外副詞、定詞、量詞也有類似問題。
 - (2) 動詞、名詞是個開放式詞集，隨時都有新詞彙產生。
 - (3) 字串結構複雜(中插結構、合併結構)。

1. 分詞標準(見附錄一：細分原則)

(1) 基本原則：理論性原則、指導原則

a. 語意無法由組成分直接相加而得到之字串應該合為一分詞單位(合併原則)。

如：飛黃騰達(成語)、撞期、吃醋(動詞組)、或多或少(副詞片語)、十二萬分(定量結構)、五月(定名結構)、三樓(定名結構)、談談(重疊結構，表嘗試)、「坐坐」就走(重疊結構，含短暫貌)、辛辛苦苦(重疊結構，表程度加強)、片片、一片片(重疊結構，具泛指意涵)、「好好」孝順父母(重疊結構，表盡力)

b. 詞類無法由組成分直接得到，應該合為一分詞單位。(合併原則)

(a) 該字串的語法功能不符合組合結果。如：動作及物動詞(喝、吃、聽)，如果前面加了「好」字，則無法後接賓語，成為不及物動詞，而且能被「很、十分、非常」等程度副詞修飾，與原來的語法成分不同。

(b) 該字串的內部結構不符合語法規律。如：游(不及物動詞)/水=在水裡游

(2) 輔助原則：操作性原則

a. 有明顯分隔標記應該切分之。(切分原則)

(a) 動賓中插：洗了一個澡

(b) 述補中插：打得破、打不破

(c) 交互中插：彎下腰去、喘不過氣來

(d) 合併中插：

動詞—上、下課

名詞—父、母親，高中、職，中山南、北路

定量—本(二)月，七、八月，1995、6年，三到四月

(e) 外來語：BBS站、user們、txt檔

b. 附著語素盡量和前後詞合為一分詞單位。(合併原則)

如：員、化等等詞綴。另外，「的、地、之」等無法作為詞綴用途，原因有二：一者其所附著之詞並無詞類限制，二者其常與詞組結合，此與詞綴的衍生方式不同，是故此三詞與前後詞一律分開。

c. 使用頻率高或共現率高的字串盡量視為一個分詞單位。(合併原則)

動詞—並列結構：進出、收放

偏正結構：大笑、改稱

動賓結構：關門、洗衣、卸貨

名詞—並列結構：春夏秋冬、輕重緩急、男女、花草

偏正結構：象牙

副詞—並列結構：暫不、既已、不再

d. 雙音節結構之偏正式動詞盡量視為一個分詞單位。(合併原則)

e. 雙音節加單音節之偏正式名詞盡量視為一個分詞單位。(合併原則)

如：線、權、車、點。防衛線、捷運線、木柵線、平均線；監護權、領導權、使用權、發言權、優先權；垃圾車、交通車、宣傳車、娃娃車；著眼點、立足點、共同點、爭議點。

f. 內部結構複雜之詞盡量切分之。(切分原則—暫行原則)

- (a) 詞組帶接詞尾：太空 計劃 室、塑膠 製品 業
- (b) 動詞帶雙音節結果補語：看 清楚、討論 完畢
- (c) 專有名詞：專名帶普名：二二八 事件、永新 加油站
詞組或句子之專名：書名、歌曲名、戲劇名
複雜結構：省 自來水 公司
- (d) 正反問句：喜歡 不 喜歡、參加 不 參加
- (e) 動賓結構、述補結構之動詞帶詞綴者：寫信 給、分紅 給

2. 分詞標準的層次劃分：信、達、雅

	應用的區別	自動分詞困難程度	建議
信	基本資料交換，取得詞象基本資料。	解決歧義切分	應於未來建立標準詞集及特殊領域詞集。
達	一般自然語言處理，建立詞雙連語言模型等。	可以用規律解決複合詞，較無複雜切分問題。	達級應將不同類型的切分與否設為標準中的可變參數。
雅	詞切分的理想境界語法，語意的抽取。如：語音合成、語意分析。	不同的語感或觀點可能造成不同切分結果，達到圓滿一致，極為困難。	除切合詞外應有標示詞性，詞內部結構的標準。雅級應探求可供客觀評估的或處理的輔助原則。

3. 搜文解字：

(1) 定量式¹的切分原則：原則上分別當作獨立的切分單位，下列情況以構詞律組合

- a. 數詞合為一分詞單位：如一萬三千
- b. 與特指詞「第」組合者：如第四十九
- c. 由定量式組合而成的時間與地方成分者：如一九九五年、三樓
- d. 量詞重疊出現，有表數量多的意思，或表普遍量化者：如「重重考驗」。

(2) 重疊式的切分原則

- a. 不是由重疊規律產生，但具有重疊形式者，合。
- b. 由重疊規律產生的 AA、ABAB、AABB 型，且 A、AB 為單一詞者，若組合後意義起化，則視為一詞，反之，若語義沒有變化，則應予以切分。
- c. 由詞組產生或有中插情形者，依據分詞助原則內部結構複雜之成分應切分。

	適用	例句
AA	單字的重疊	圈圈、問問
AAB	重賓式複合詞	見見面、聊聊天
	單字動詞重疊後+「看」	試試看、拼拼看
ABB	成詞性的形容動詞	孤零零、靜悄悄
	擬聲詞	淅哩哩、嘩啦啦
	單字動詞+「看看」	加看看、穿看看
AABB	狀態動詞	庸庸碌碌、歡歡喜喜
	單字詞重疊並列	打打殺殺、大大小小
	少數定詞	許許多多
ABAB	動作動詞	考量考量、通融通融
	狀態性動詞	悠哉悠哉、涼快涼快
	雙音節擬聲詞	轟隆轟隆、喀拉喀拉
	定量式重疊	一聲一聲、一勺一勺
A 裡 AB	負面性評價的形容性動詞	傻裡傻氣、糊裡糊塗
AXA	單音節動詞重疊，可中插副詞「一、了」	看一看、數了數

¹ 定量式複合詞：定詞與量詞的組合性極高，其搭配由構詞律產生，可以無限制的衍生，且合成的成分保留了原有的意義，乃介於詞與詞組之間的過渡詞。此複合詞在中文語料庫中多被切分，因其切分可以反映漢語「量詞」的特點，加上量詞有時可以單獨成句，數詞與量詞間也可以加入修飾成分，並且可以避免過度冗長的缺失。

(3) 詞綴、接頭/接尾的切分原則

- a. 詞綴²：(a) 衍生詞綴(derivational affix)：衍生性低者，收入詞典。
(b) 屈折詞綴(inflexional affix)：有獨立固定的語法功能，所以可以獨立使用，且不影響緊鄰的語法類型，所以一律切分。如：了、著、過、們、得、看、看看。
(c) 關係子句標誌(relative clause marker)：音前方組成成分為詞組形式，是故切分之。如：者、的。
前詞綴：老、小、第、阿、可、好、互、相、自
後詞綴：氏、某、化、度、性、家、長、師、員、兒

(4) 述補式的切分原則：詞類無法由組成成分直接得到，應該合為一分詞單位。(合併原則)
述補式複合詞：由兩個據謂語性的詞素所組成，第二個述詞(predicate)乃描述第一個述詞所發生的結果，兩者互補卻又各自獨立。

合詞：喝光、吹乾、打歪、踢破、穿破

分詞：第二部分較為複雜時，則依輔 6 原則切分之。如：走出來、跑上去、丟進去、蔓延開來。
第二部分有中插時，因此中插成分為語法詞綴，故切分之。如：不、了。

(5) 動詞後接「於」的切分原則

分 a.動詞和「於」結合成詞，但後接時間和地方成分例外。

b.若「於」表示「在」、「由於」則予以切分。

合 a.前接動詞無法單獨使用，表示「動詞+於」已成為新詞。

b.和動詞結合後，意義不同於動詞原來的意思，表示「動詞+於」已成為新詞。

a)後接時間和地點的介詞。如：生於日據時代、建於清朝。

b)後接動作終點。如：仁在實踐上表現於忠恕之道。

c)動作+於+施事者。如：舊有教室在民國六十年毀於娜定颱風。

d)動詞後的「於」，可以省略。

e)動詞與「於」結合緊密，不可切分。如：批評不 等於 鬥爭、昧於 私利。

(6) 動詞後接「為/成/作」的切分原則

(7) 動詞後接「給」的切分原則：除了在述補和動賓結構予以切分外，其餘都採取動詞和給結合成詞之方案。

(8) 動詞後接「有」的切分原則：一律視為一切分單位。

a.動詞+「有」多半會引起語法變化，具有地方詞倒裝的現象。

b.動作動詞後方加「有」，意義會轉變為狀態動詞。

c.「V-有」無法被「沒」否定，已經失去「有」的特性。

(9) 「的、地、之」的切分原則

a. 的、地：視為一分詞單位，與醜、後成分區分開來。

b. 之，與後面成分合詞，成為方位詞(如之餘)或形容性動詞後的修飾語(如之至、之極)收列在辭典中，或是與數字合為分數或序數，由構詞律結合為一分詞單位。

A.的

a.修飾語+的+名詞：美麗的姑娘。

b.修飾語+的+動詞(包括形容詞動詞)：意外的碰見。

c.動詞+的+結果或程度的補語：來的巧。

d.句子+的：他打從新竹來的。

B.地：有些「的」可以改為「地」

C.之

a.代名詞

b.修飾語+之+名詞

c.修飾語+之+動詞

² a.衍生詞綴是以構詞律結合的「構詞詞綴」(morphological affix)，b.c.則是「語法詞綴」(grammatical affix)。

(10) 否定式的切分原則：不、沒、別、休、非

- a. 不：
- (a) 當「不」出現於動詞前：應與動詞切分。但與其結合之動詞結合後並無肯定之對應詞或失去語意結合性，則予以合併。如：不 說謊、不 用功；不力、不禁、不對勁、不實、不休、不當；不錯、不斷、不明、不平、不快、不齒。
 - (b) 當「不」出現於法相詞前：相結合標註「副詞」。如：不可能、不應該、不得、不肖。
 - (c) 當「不」出現於副詞前，若副詞為一獨立單位，則宜分詞：不 常、不 一定。但與其結合之動詞結合後並無肯定之對應詞，或失去語意結合性，或切分後會留下一附著成分，或出現頻率高時，則合之：如：不免、不外、不復、不遑、不料、不再、不自禁。
 - (d) 當「不」出現於名詞詞頭，若此名詞為一語意單位，則應合之：不是、不同、不歸路。
 - (e) 當「不」出現於成詞中：不安於室、不可一世、不勞而獲、不虛此行。
 - (f) 當「不」出現於述補式的否定可能式中：打 不 死、走 不 動、看 不 見。
 - (g) 當「不」出現於正-反問句形式(A-not-A)中：可不可以，能不能，亮 不 亮，喜歡 不 喜歡。
- b. 沒有：「有」表示了三種語意--完成、擁有、存在，「沒」以否定「有」，成互補關係，所以「沒有」一律合詞。
- c. 非：作為名詞前綴則合之，作為動詞、副詞則單獨成詞。
- d. 休、別：在祈使句中，可單獨成詞，標註為副詞。

(11) 正反問句(A-not-A 問句)的切分原則

- a. 合：附著語素應盡量和前後詞合為一分詞單位
- b. 分：重疊成分不完整，如：認不認識、喜不喜歡、噁不噁心、低不低級、幽不幽默。動賓結構中重疊成分不完整，或有中插時；如：開 不 開 他 的 玩 笑。

(12) 中插詞的切分原則：所有的中插成分皆分詞，前部分(+p1)+後部分(+p2)

動詞	動補中插	中插表能力的「得」、「不」	打的破、打不破
	動賓中插	中插補語(某些補語可加「得」、「不」)	洗好澡、幫得上忙
		中插時態標記	洗了澡、幫過忙
		中插名詞修飾語成領格	幫他的忙、洗個痛快澡
		中插補語+時態標記	洗好了澡、幫上了忙
		中插時態+名詞修飾語成領格	幫過他的忙
		中插補語+名詞修飾語成領格(某些補語可加「得」、「不」)	幫得上他的忙
		中插補語+時態標記+名詞修飾語成領格	幫上了他的忙
	動補動賓 交互中插	動補中插賓語	彎下 腰 去
		動賓中插補語	彎 下 腰 去
		動補中插賓語	喘不過 氣 來
		動賓中插補語(補語前可加「得」、「不」)	喘 不 過 氣 來
	重疊中插	中插「了」	談了談、笑了笑
		中插「一」	談一談、笑一笑
中插「了」、「一」		談了一談、笑了一笑	
名詞	合併中插	中插標點	父、母親，中山南、北路
		中插連接詞	母或父親

(13) 合併形式的切分原則：無中插時，一律合併。

- a. 詞頭：國內外、高中職
- b. 詞尾：青少年
- c. 頭尾合併：中山南北路
- d. 套裝合併：環保署長
 - (a) 不含行政區域的單純套裝予以合併：建設局長、保險業者
 - (b) 含地名、行政區域的單純套裝予以切分：台北 市長、高雄 縣長
 - (c) 複雜套裝予以切分：私貨 處理 股長

(14) 動詞後修飾語的切分原則：與動詞切分開來

文言語法：不下、不力、不休、不羈、不成、不誤、不已、有方、有加、有成、有序、有案、有素、無方、無力、無功、無比、無門、無度、無常、若渴、若狂、若此、若常、若鶩

(15) 專有名稱的切分原則

- a. 單純詞：愛因斯坦、穆罕默德
- b. 縮寫：北二高、奧運會
- c. 頭尾合併詞：忠孝東西路
- d. 專名+普名：合一普名是接尾詞，如：阿美族、陳董、桃園廠、高雄港、竹聯幫、山口組
分一普名是自由詞素，二二八 事變、平漢 鐵路
- e. 複雜詞：分
- f. 詞組或句子：分

(16) 成語諺語的切分原則

成語作為一分詞單位之因素：

1. 語意無法由組合語素(morpheme)得知。
2. 內部結構複雜似可切分，然語序固定，格遵詞語自主律(lexical integrity)，應該和複合詞相同在詞彙中處理。
3. 言簡易該，一旦予以切分，如此會產生無法處理的附著語素。

附錄一：細分原則

			合例／分例	
複合詞	定量式複合詞	數詞	合例：一千八百、百分之三十、三十%、三又二分之一、六十六點五、五成三。	
		表時間、地點	合例：一九九五年 三月 六日 二點 二十分	
		普通定量詞	三位、五十二隻、三又二分之一打、七十餘位、七十位餘	
	複合動詞	並列式	合例：發交相關單位	
		偏正式	合例：增蓋、冷燻、勇奪	
		動賓式	合例：洗澡、回國	
		主謂式	合例：頭昏	
		動補式		
			方向補語	合例：跑上來
			結果補語補語是單音節	合例：打壞
		補語是雙音節	合例：打掃 乾淨	
	V-給、V-到、V-於、	V-給	合，但動詞本身是動賓或動補結構則分：批發給、寫信給、分紅給、取出給、退回去給	
	V-有、V-爲、V-成、	V-到：	合，但後接時間，補語和數量詞時分：接觸到、聊到半夜、走到腿酸、加到兩百萬	
	V-作 另行規定	V-於	分；但動詞是附著詞、合併後意義改變、表示比較，則合：生於台北、吝於、有感於、大於、優越於	
		V-有	合	
		V-爲	合	
		V-成	合	
		V-作	合	
	偏正式複	簡單式	合	

	合名詞		
		帶詞綴	修路費、準女婿
		語意無組合性	土包子、鐵公雞
		專指	白菜、黑板
		不能中插	太陽眼鏡、旅行支票
		使用頻率高	牛肉麵
		含有附著成分	奇案、勇將
		複雜式	
		簡短常見式	合 例：借書證、租車費
		冗長、少見式	分 例：欲 偷渡 到 美國 者
	複合副詞		合例：竟是、並非、暫不、既已…
	名方式複 合詞	語意失去組合性	合 例：手上、腦中、眼裡
		有清楚的指涉	合 例：目前、月底
	定語+ 的、狀語+ 的/地		分 例：詞+的/地 我的書、代理的人員 詞組+的/地 很神氣的天鵝、聲音平緩的讀著稿
	簡稱		合 例：男單、女網、空姐、影視、化工、音像
	合併詞	無中插	詞頭合併 合 例：國內外、高中職
		詞尾合併	合 例：父母親、公私立
		頭尾合併	合 例：中山南北路
		套裝合併	前面是專名 分 例：台北 市長、新竹 縣政府 前面是其他 合 例：事務局長、體育司長
		有中插	分 例：初、高中
	重疊詞：合	動詞	嘗試貌：談談、想想、研究研究、說說看 暫時貌：坐坐 就 走、擦擦 就 可 程度貌：胖胖 的、辛辛苦苦
		名詞	車車、狗狗、痘痘
		量詞	片片、一片片（但：一片一片、一片又一片）
		擬聲詞	叮叮噹噹、乒乒乓乓（但：嘩啦 嘩啦、哈哈 哈哈）
	專有名詞	單純詞	合 例：胡適、布農、貝多芬、阿爾及利亞、宇宙光
		專名+普名：普名 是詞綴	合 例：阿美族、光復橋、竹聯幫、桃園廠、王董
		普名是自由詞素	分 例：胡 先生、二二八 事變、永新 加油站
		縮寫	合 例：勞基法、奧申委、文建會、臺三線、北二高
		複雜詞	分 例：第一 信用 合作社、省 自來水 公司
		詞組或句子	分 例：前世 今生、鯨魚 的 生 與 死
	中插詞	分 動賓中插	分 例：洗 了 一 個 澡
		動補中插	分 例：打 得 破、打 不 破
		重疊中插	分 例：笑 了 笑、哭 一 哭
	A-not-A		分：喜 歡 不 喜 歡 喜 不 喜 歡

附錄二：

		依據原則		
動詞	並列結構	合	基 12 輔 23	讀誦文章、擴建完畢
		分	不符基 12 輔 23	叮 咬不停
	偏正結構	合	基 12 輔 234	改祭瓜果、大笑不已、高奏凱歌
		分		
	主謂結構	合	基 12 輔 23	
		分		價 升 量 減
	動賓結構	合	基 12 輔 23	騙人、關門、洗衣、拔草、卸貨
		分	中插，輔 16	騙了人、洗了一個澡
	述補結構	合	基 12	到(接觸到)、為(批評為)、成(剪成)、作(換作)
		分	中插，輔 1	
			結果補語為雙音節，輔 6	
	動補中插	分	輔 1	打得破、打不破
	重疊結構	合	基 1，中插：輔 1	
	帶詞綴	合	輔 2	送給、批發給、收有
		分	動賓結構、述補結構	
正反問句結構	合	不完整形式：輔 2	喜不喜歡、盜不盜壘、開不開刀	
	分	完整形式：輔 1		
		中插：輔 6	開 不 開 他 玩 笑	
合併結構	合	基 1	上下學、出入境	
	分	中插：輔 6	上、下學，出、入境	
中插結構	分	輔 1		
普通名詞	並列結構	合	基 12 輔 23	春夏秋冬、輕重緩急、男女、花草
	偏正結構	分	輔 6	公職人員、財務報表、公共設施
		合	基 12 輔 23	大雨、象牙
	帶衍生詞綴、接頭/接尾詞	分	輔 6	太空計劃室、國際關係組
		合	輔 25	電腦室、業務部
	簡稱	合	基 1	男單、女網、空姐、影視、化工、音像
合併結構	合	基 1，帶專名則否	詞頭合併：高中職、國內外 詞尾合併：父母親、公私立 套裝合併：事務局長、台北市長	
中插結構	分	輔 1	春、夏、秋、冬，男、女、老、少	
專有名詞	單純詞		基 1 合，輔 6 分	胡適、桂林、貝多芬、阿爾及利亞
	專名+普名		基 1 合，輔 6 分	普名是接尾詞：阿美族、光復橋、竹聯幫 普名是自由語素：胡先生、平漢鐵路
	縮寫		基 1 合，輔 6 分	勞基法、奧申委、文建會、台三線
	複雜詞	分	輔 6	台北市第一信用合作社
	詞組或句子		基 1 合，輔 6 分	那一年我們都很酷
定量式	定詞	分	為數字一律合併	三十五、三又六分之一、20%
	量詞	分	重疊結構則合	片片、個個
	定量詞	合	重疊結構、表時地	一片片、八十四年、七十巷
分			一片、一個	
副詞		合	基 12、輔 23、重疊	暫不、既已、不料、不變、偷偷
		分		砰砰、咻咻咻
成語諺語		合	基 1	陰錯陽差、一不做二不休、一而再再而三
		分	輔 6	話不投機半句多、虎落平陽被犬欺

附錄三：語法詞綴、衍生詞綴、接頭/接尾詞

語法後綴	名詞後綴	們
	動詞後綴	了、著、過、看、看看
	關係子句標記	者、的/地
衍生前綴	名詞前綴	老、小、第、阿
	動詞前綴	可、好、相、互、自
衍生後綴	名詞後綴	氏、某、度、性、家、長、師、員
	動詞後綴	化
	多功能後綴	兒、儿
接頭詞	名詞	副、非、多、高、低、超、零、單、雙、前、準、總、主、代、泛
	動詞	加、以、重、增、轉、搶、合、代、偷、誤、補、預、借、回、速、 複、超、試、抽、趕、共、大、小、對、耐、續
接尾詞	動詞	給、有
	非謂形容詞	用
	名詞	壩、波、簿、盃、碑、杯、輩、包、堡、報、班、版、板、本、幫、 榜、棒、別、錶、表、濱、病、比、壁、幣、埔、部、兵
		坡、婆、排、牌、派、盤、盆、旁、棚、票、篇、片、品、瓶、坪、 癖、鋪、舖
		碼、膜、饅、模、沫、脈、帽、盟、苗、廟、面、皿、迷、木、目
		法、費、販、犯、粉、坊、方、房、風、峰、符、服
		地、帶、袋、島、道、豆、單、彈、檔、黨、調、點、殿、店、町、 堆、隊、端、段、洞、凍
		臺、台、態、陶、套、頭、攤、灘、壇、潭、罈、堂、帖、廳、亭、 艇、題、體、團、通、筒、桶、徒、圖、土
		年、能、泥、農、女
		類、佬、樓、露、欄、郎、倆、列、寮、料、瘤、連、戀、林、量、 令、嶺、里、禮、吏、力、輪、論、爐、路、錄、律、率
		閣、糕、膏、稿、鎬、溝、乾、桿、感、缸、岡、崗、港、卦、鍋、 國、果、粿、鬼、櫃、觀、關、官、管、館、罐、功、宮、工、公、 骨、股、鼓、谷
		卡、科、客、課、考、坑、窟、塊、款、狂、孔、褲、庫
		盒、河、號、函、漢、行、會、花、話、貨、環、荒、鬍、壺、糊、 戶
		機、肌、夾、岬、假、價、架、街、界、節、結、膠、間、鍵、艦、 金、江、獎、醬、匠、經、精、鏡、級、籍、集、計、劑、記、祭、 季、圈、捲、君、軍、郡、局、劇、卷
		期、橋、球、琴、腔、槍、旗、氣、器、區、曲、拳、券、權、群、 裙
		溪、峽、俠、匣、鞋、屑、學、秀、嫌、險、線、縣、腺、人、心、 相、箱、鄉、項、巷、星、型、形、姓、席、系、穴、軒、訊

		渣、宅、齋、寨、罩、照、州、洲、展、站、戰、針、鎮、陣、章、 症、證、証、珠、株、桌、傳、庄、莊、莊、狀、主、史、冢、種、 塚、竹、拄、質、值、職、誌、製、制
		痴、癡、差、潮、禪、廠、場、城、川、串、窗、處、池
		生、石、舍、社、衫、山、扇、身、省、商、聲、書、刷、數、稅、 霜、署、術、樹、束、市、室、式
		熱、日、仔、澡、作、座、罪、宗、族、組、子、字
		側、冊、財、操、槽、餐、艙、層、村、叢、詞、祠
		絲、司、色、賽、傘、散、酥、所、鬆、素、寺
		額、癌、案、衣、液、頁、業、藥、炎、岩、巖、堰、宴、印、營、 儀、椅、尾、裔
		屋、窩、位、味、灣、彎、丸、紋、文、王、網、舞、物
		源、園、苑、語、嶼、欲、慾、浴、域、樂