

中國國家標準	<h1>中文資訊處理分詞規範</h1>	總號	
<h1>CNS</h1>		類號	X

SEGMENTATION PRINCIPLE FOR CHINESE LANGUAGE PROCESSING

1. 適用範圍：本標準規定資訊處理用分詞標準研擬草案之原則、層次劃分及應用實例，並且經後處理之後適用於中文資訊檢索、機器翻譯和文句校正等。
2. 用語釋義
 - (1) 分詞標準：根據語言學的資訊處理為著眼點，規範中文字串基本語意單位切分的遵循標準。
 - (2) 定義：用來說明中文分詞的基本單位。
 - (3) 基本原則：中文分詞標準的一般性原則，從語意、語法兩方面來規範，符合語言學理論。
 - (4) 輔助原則：中文分詞的操作型原則，輔助原則可因需要而有變異性。
 - (5) 層次劃分：依自動化處理的難易，規定不同階段應達成的分詞目標。
 - (6) 信級：在本規範的前提下，所定的的操作原則，依標準詞典將詞切分，用來作基本資料交換。
 - (7) 達級：在本規範的前提下，所定的的操作原則，是簡單構詞規律所能組合成詞的層次，用於大部分自然語言處理。
 - (8) 雅級：在本規範的前提下，所定的的操作原則，是符合語言學理論的理想階層，可用於剖析。
 - (9) 操作型定義：在本規範的前提下，依詞類、結構等，列出各類型詞語，供使用者實際操作時的依據。
 - (10) 標準辭典：是理想中的辭典，辭典收納的詞能符合分詞標準，並能與時俱增，隨時更新編入語言演變產生的新詞。
 - (11) 附著語素：有獨立意義卻無法獨立扮演一個語法功能的語言成分。
 - (12) 衍生詞綴：具有衍生性的附著語素，可由構詞律組成複合詞，對資訊處理的困難度不大，在達級即可達成組合詞的目標。
 - (13) 語法詞綴：具有固定獨立的語法功能，且不影響緊鄰成分語法類別的詞綴，包括「了、著、過、看、看看、們、者」等。
 - (14) 接頭詞：附加於別的成分之前構成複合詞的詞，在雅級處理其組合。如「準一、多一、非一」。
 - (15) 接尾詞：附加於別的成分之後構成複合詞的詞，在雅級處理其組合。如「一盃、一盒、一觀」。

第二次修訂：87.05.29

(共 27 頁)

公 布 日 期 年 月 日	<h2>經濟部中央標準局印行</h2>	修 訂 日 期 年 月 日
------------------	---------------------	------------------

- (16) 並列結構：具有相同詞類的成分並排出現，如兩個動詞並列「研究」或兩個名詞並列「兄弟」。
- (17) 偏正結構：修飾成分與主要語形成的結構，如「汽車」是偏正式名詞，「微笑」是偏正式動詞。
- (18) 主謂結構：主語性及謂語性成分所構成的結構，例如：臉紅、心跳。
- (19) 動賓結構：動詞與受詞結合形成的結構，例如：開門、打嗝。
- (20) 述補結構：由動詞或形容詞與描述其動作結果或狀態的詞組合而成，例如：走光、洗掉。

3. 分詞原則及層次劃分

3.1 分詞原則：規定分詞之依據，包含定義、兩條基本原則、六條輔助原則。

3.1.1 定義：具有獨立意義，且扮演固定詞類的字串視為一分詞單位。

3.1.2 基本原則

(1) 語意無法由組成成分直接相加而得到之字串應該合為一分詞單位。

例 1. 下列字串因其組合後語意改變皆應視為一個分詞單位：飛黃騰達（成語），撞期、吃醋（熟語），或多或少、十二萬分（定量結構），五月（定名結構：不是五個月）、三樓（定名結構：不是三層樓），談談（重疊結構：表嘗試）、坐坐（重疊結構：短暫貌）、辛辛苦苦（重疊結構：程度加強）、片片（重疊結構：具泛指意涵）⁽¹⁾。

註⁽¹⁾ 當重疊結構之意義未失組合性，則不予合併。例如“坐 坐 坐、哈 哈、叮 噹 叮噹”不須組合成一個詞，因該字串之語意可從每個成分組合而成，並無多出的詞意。

例 2. 合併結構，像是“上下課、高中職、中山南北路”，依此原則也應該合併為一個詞。因為該字串的意義並非“上”加“下課”、“高中”加“職”，“中山南”加“北路”，而是“上課”加“下課”、“高中”加“高職”、“中山南路”加“中山北路”，可見合併結構的意義不等於組合意義，故應合併⁽²⁾。

註⁽²⁾ 唯帶專名之合併詞，像是“台北市長”（“台北市”加“市長”）、“新竹縣政府”（“新竹縣”加“縣政府”），因切分後前方的專名和後方的名詞皆可獨用，意義可以組合成，故仍予以切分。

(2) 詞類無法由組成成分直接得到，應該合為一分詞單位。

例 1. 動作及物動詞“喝、吃、聽”前面加“好”構成“好喝、好吃、好聽”，不能再加賓語，成為不及物動詞，且能被程度副詞“很、十分、非常”修飾，與原來的語法特性不同，故可視為一個分詞成分。

例 2. “那隻狗不會游水”中“游水”指的是“在水裡游”，但“游”是不及物動詞，不可直接後接名詞。因此，“游水”不符合動詞“游”的語法規律，故應合併之。

備考：基本原則(1)(2)配合定義分詞單位，視同選詞之標準，故為合併原則。

3.1.3. 輔助原則

(1) 有明顯分隔標記應該切分之。

例 1. 動賓中插：洗了一個澡

例 2. 述補中插：打得破、打不破⁽³⁾

例 3. 交互中插：彎下腰去、喘不過氣來

例 4. 合併中插：動詞：上、下課

例 5. 合併中插：名詞：父、母親，高中、職，中山南、北路

例 6. 合併中插：定量：本(二)月，七、八月，1995、6年，三到四月⁽⁴⁾

例 7. 外來語：BBS 站⁽⁵⁾

備考：中插的分隔標記可能是詞、標點符號或是外來語。

註⁽³⁾ 但像“養得起、養不起”、“處得來、處不來”因無相對應之“養起”、“處來”，所以視為一分詞單位，不予切分。

註⁽⁴⁾ 數詞及表時間、地點或編號之詞雖含有標點符號，但是這些符號不具標點符號功能，所以不算是中插，仍維持合併。

例：七、五00；三·六；2/28（二月二十八號）；3:30（三點三十分）；二0~一號（門牌號碼）；AB-8888（車牌號碼）

註⁽⁵⁾ 有一些包含外來語的詞，如“卡拉 OK、K 書”已為大眾普遍使用，而且這些詞語常常一起出現，因此不予以切分，維持合併。

(2) 附著語素盡量和前後詞合為一個分詞單位。

例 1. “情勢立告逆轉”中的“立”是附著語素，表“立刻”之意，我們依此原則將“立告”合為一個偏正式複合動詞。

例 2. “吝”也是個附著語素，多半出現在“吝嗇”“吝惜”中，但依此原則“不吝”“吝於”也會被合併成一個動詞。

例 3. 詞綴也是附著語素，因此帶詞綴之字串也應合為一詞。例如：“演員、救生員、隊員、查哨員、技術員、組成員、督導員、郵務員…”“現代化、合理化、泛政治化、民營化、地下化、本土化、小丑化、多元化…”⁽⁶⁾⁽⁷⁾。

註⁽⁶⁾ 這些詞在詞典中收不勝收，必須藉構詞律由電腦自動結合成詞。但是從電腦處理的角度來看，在初步的處理時並不容易達成自動合詞的目標，必須依不同層次分階段達成，因此依附著詞結合難易的程度分為詞綴及接頭/接尾詞，挑選出衍生性強的詞綴及接頭/接尾詞作為分詞的參考依據，請參照附錄 2。

註⁽⁷⁾ “的、地、之”雖通常被視為詞綴，但是由於下列兩個理由我們不將它們當作詞綴處理。一、它們所附著之詞幹無詞類限制，無論名詞、動詞、副詞、數量詞甚至句子皆能帶這些詞綴，這和一般詞綴表現不一；二、它們常和詞組結合，如“常常和官員打交道的記者”“欲退出選委會之人”，這點也和一般詞綴的衍生方式不同，所以這三個詞將和前後詞一律斷開。

備考：有時可能遇到附著語素無法和前後詞合成一個語言成分的情況，如“爲什麼還吝而不做呢？”，在此情形下只好將附著詞“吝”斷開，依其在該句中所扮演的功能給予詞類。

(3) 使用頻率高或共現率高的字串盡量視爲一個分詞單位。

例 1. 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、拔草、卸貨、……

例 2. 名詞：

並列結構：春夏秋冬、輕重緩急、男女、花草、……

偏正結構：象牙、……

例 3. 副詞：暫不、既已、不再、……

備考：1.這些字串因爲常常一起出現，所以其結合較緊密，較少見中插情形。縱使這些字串的語意、語法功能未失組合性、也不含附著語素，仍可因此原則合爲一個詞。

2.這條原則的難處，在於如何得出使用頻率，以及區分值應該設在何處，對於一些字串此原則是否適用就成了見仁見智的情形，因此需要一部標準辭典作爲區分詞和非詞的依據。故這條原則爲一條選用原則。

(4) 雙音節結構之偏正式動詞盡量視爲一個分詞單位⁽⁸⁾。

例 1. “緊追其後”中的“緊追”語意、語法功能雖然未失組合性，不含附著語素，也不是常見字串，仍可依此原則合併之。

註⁽⁸⁾ 此原則並不用於動賓及主謂式複合動詞。所以“警察無故擒人”“股市陷入價升量減的走勢”中“擒人”和“價升量減”不會因此原則合併。

(5) 雙音節加單音節之偏正式名詞盡量視爲一個分詞單位。

例 1. “線”所構成的成分“防衛線、捷運線、木柵線、平均線”。

例 2. “權”所構成的成分“監護權、領導權、使用權、發言權、優先權”。

例 3. “車”所構成的成分“垃圾車、交通車、宣傳車、娃娃車”。

例 4. “點”所構成的成分“著眼點、立足點、共同點、爭議點”。

備考：例中所舉的單音節名詞本身可獨立成詞，但是常與前面的雙音節成分結合緊密。故可合爲一分詞單位。從與其他成分結合的觀點來看，這些單音節名詞也可視爲接尾詞，與衍生性附著語素並列在接尾詞之列（參照附錄 2）。

(6) 內部結構複雜之詞盡量切分之。

例 1. 詞組帶接尾詞：太空 計劃 室、塑膠 製品 業

例 2. 動詞帶雙音節結果補語：看 清楚、討論 完畢

例 3. 專有名稱

專名帶普名：胡 先生、平漢 鐵路、二二八 事變、永新 加油站

詞組或句子之專名：鯨魚的生與死（書名）、那一年我們都很酷（戲劇名）

複雜結構：省自來水公司、台北市第一信用合作社、輔大景觀設計系、中文分詞規範研究計畫

例 4. 正反問句：喜歡不喜歡、相信不相信

例 5. 動賓結構、述補結構之動詞帶“給”：分紅給、取出給、退回去給

備考：1.上例結構依前述原則可合為一個詞，但由於其內部結構具組合性，合併起來過於冗長，故不予合併。

2.輔助原則(1)(6)為切分原則，(2)~(5)為合併原則。

3.2 層次劃分：依電腦自動化處理分詞的難易程度及實際使用情況，分信級、達級、雅級三個層次。

(1) 信級：凡是收錄在標準辭典中的詞一律斷開。

(2) 達級：能以構詞律組合出來的詞在達級合併。

(3) 雅級：無法完全收錄在辭典中的詞在雅級合併。

備考：分詞之三級標準分別有其特定應用範圍：(1)信級是文獻資料交換之基本層級，在此層次，除了可解決歧義，還符合基本資料交換的最基本形式；(2)達級用於大部份自然語言處理；(3)雅級是符合語言學理論的理想階層，也可用於剖析自然語言。但是在雅級，某些類型的詞仍可找出組合的規律，另一方面，雅級的自動化不易達到，因此不硬性規定達級和雅級間可處理的類型，而改採浮動等級。採取浮動等級更可以符合不同需求，而且以表列方式可以提供給使用者實際操作型的根據（參照附錄 3）。

附 錄 1

示 例

1. 詞類分合標準：針對不同詞性、不同結構的分合，說明所引用的分詞原則。

定義：具有獨立意義，且扮演固定詞類的字串視為一分詞單位。

基本原則：

- (1) 語意無法由組合成分直接相加而得到之字串應該合為一分詞單位。
- (2) 詞類無法由組合成分直接得到，應該合為一分詞單位。

輔助原則

- (1) 有明顯分隔標記應該切分之。
- (2) 附著語素盡量和前後詞合為一個分詞單位。
- (3) 使用頻率高或共現率高的字串盡量視為一個分詞單位。
- (4) 雙音節結構之偏正式動詞盡量視為一個分詞單位。
- (5) 雙音節加單音節之偏正式名詞盡量視為一個分詞單位。
- (6) 內部結構複雜之詞盡量切分之。

1.1 動詞

- (1) 並列結構：若符合基本原則(1)(2)或輔助原則(2)(3)任何一項則合併，否則予以切分。

例 1. 讀誦文章、擴建完畢 [符合輔助原則(2)]

例 2. 叮 咬 不停 [不符合上述原則]

- (2) 偏正結構：若符合基本原則(1)(2)、輔助原則(2)(3)(4)任何一項，則合併。

例 1. 改祭瓜果、大笑不已 [符合輔助原則(4)]

例 2. 高奏凱歌 [符合輔助原則(3)]

- (3) 主謂結構：若符合基本原則(1)(2)、輔助原則(2)(3)任何一項，則合併。

例：陷入價 升 量 減的走勢 [不符合任何一項]

- (4) 動賓結構：若符合基本原則(1)(2)、輔助原則(2)(3)任何一項，則合併。又，中插情形依輔(1)切分。

例 1. 關門、洗衣、拔草、卸貨 [符合輔助原則(3)]

例 2. 洗了一個澡、卸了貨 [符合輔助原則(1)(6)]

- (5) 述補結構：依基本原則(1)(2)一律合併。唯當補語是結果補語且是雙音節時，依輔助原則(6)切分。又，中插情形依輔助原則(1)切分。

例 1. 哭濕枕頭、爬上山頭、走進去、看 清楚、清洗 完畢

例 2. 到：接觸到、認知到、跑到 [是述補結構，故合併]

聊 到半夜、走 到腿酸、加 到十萬 [不是述補結構，故切分]

例 3. 為：譯為、流為、批評為、選拔為 [是述補結構，故合併]

例 4. 成：擠成、剪成、歸劃成、堆積成 [是述補結構，故合併]

- 例 5. 作：鑄作、換作、署名作、轉變作 [是述補結構，故合併]
- 例 6. 述補中插：打得破、打不破 [符合輔助原則(1)]
- (6) 重疊結構：若符合基本原則(1)，則合併。唯中插情形依輔助原則(1)切分。
- 例 1. 嘗試貌：談談、研究研究 [符合基本原則(1)]
說說 看、說 看看 [符合輔助原則(6)依定義切分]
- 例 2. 暫時貌：坐坐就走、擦擦即可 [符合基本原則(1)]
- 例 3. 程度貌：胖胖的、辛辛苦苦、慢吞吞 [符合基本原則(1)]
- 例 4. 其他：坐 坐 坐 [不符合任何一項]
- 例 5. 重疊中插：笑 了 笑、哭 一 哭 [符合輔助原則(1)]
- (7) 帶接尾詞：依輔助原則(2)應合併。唯當動詞詞幹是動賓、述補結構依輔助原則(6)切分。
- 例 1. 送給、贈送給、批發給 [符合輔助原則(2)]
- 例 2. 寫信 給、分紅 給、取出 給、退回去 給 [依輔助原則(6)切分]
- 例 3. 收有、列印有 [符合輔助原則(2)]
- (8) 正反問句結構：完整形式依輔助原則(6)將之切分，不完整形式則依基本原則(1)、輔助原則(2)合併。
- 例 1：喜歡 不 喜歡
- 例 2：喜不喜
- (9) 合併結構：依基本原則(1)應合併。唯中插時依輔助原則(1)切分。
- 例 1. 上下學，入出境
- 例 2. 上、下課，入、出境
- (10) 中插結構：依輔助原則(1)必須切分。
- 例：動賓、述補交互中插：幫 得 上 忙、喘 不 過 氣 來

1.2. 普通名詞

- (1) 並列結構：若符合基本原則(1)(2)、輔助原則(2)(3)任何一項，則合併。
- 例：春夏秋冬、輕重緩急、男女、花草 [符合輔助原則(3)]
- (2) 偏正結構：若符合基本原則(1)(2)、輔助原則(2)(3)任何一項，則合併。
- 例：象牙 [符合輔助原則(3)]
- 備考：“公職 人員、財務 報表、公共 設施”依輔助原則(6)予以切分。
- (3) 重疊結構：依基本原則(1)應合併。
- 例：一隻狗狗、長痘痘、小車車
- (4) 帶衍生詞綴、接頭/接尾詞：依輔助原則(2)、(5)應合併。唯當詞組帶詞綴時，依輔助原則(6)應切分。
- 例 1. 電腦室、業務部
- 例 2. 太空 計畫 室、國際 關係 組

(5) 簡稱：依基本原則(1)應合併。

例：男單、女網、空姐、影視、化工、音像

(6) 合併結構：依基本原則(1)應合併。唯帶專名之合併結構不符合基本原則(1)不需合併。

例 1. 詞頭合併：高中職、國內外

例 2. 詞尾合併：父母親、公私立

例 3. 套裝合併：事務局長、台北市長、新竹縣政府

(7) 中插結構：依輔助原則(1)應切合。

例：並列中插：春、夏、秋、冬、男、女、老、少

1.3 專有名詞：依基本原則(1)應一律合併。唯依輔助原則(6)有幾種結構複雜之專有名詞將不予合併。

例 1. 單純詞：胡適、桂林、布農、貝多芬、克寧、阿爾及利亞

例 2. 專名+普名（普名是接尾詞）：阿美族、光復橋、竹聯幫

例 3. 專名+普名（普名是自由語素）：胡先生、平漢鐵路、二二八事變

例 4. 縮寫：勞基法、奧申委、文建會、台三線、中常會

例 5. 複雜詞：台北市第一信用合作社、省自來水公司

例 6. 詞組或句子：鯨魚的生與死（書名）、那一年我們都很酷（戲劇名）

1.4 定量式

(1) 定詞：依分詞定義應予以切分。唯數詞依基本原則(1)一律合併。

例：三十五、八萬零二十點七、三又二分之一、百分之四十、三八.000、2.3、2.0%

(2) 量詞：依分詞定義應予以切分。唯重疊結構依基本原則(1)一律合併。

例：片片、個個

(3) 定量詞：依基本分詞原則定詞和量詞應切分。唯重疊結構依基本原則(1)則予以合併，又表時間、地點之定量詞依基本原則(1)應合併。

例 1. 二片、二個 [依定義切分]

例 2. 一片片、一個個 [符合基本原則(1)具泛指功能]

例 3. 二片二片、二個二個 [不符合基本原則(1)未具泛指功能]

例 4. 八十四年九月一日三時二十分 [符合基本原則(1)]

例 5. 七十巷二十號之一三樓 [符合基本原則(1)]

1.5 副詞：唯有符合基本原則(1)、(2)、輔助原則(2)、(3)任何一項才予以合併。又重疊結構若符合基本原則(1)應予以合併。

例 1. 暫不、既已 [符合輔助原則(3)]

例 2. 不過、要不是、或早或晚 [符合基本原則(1)]

例 3. 不料、不便 [符合輔助原則(2)]

例 4. 偷偷、悄悄 [符合基本原則(1)或輔助原則(2)]

例 5. 叮噓 叮噓、砰 砰、咻 咻 咻 [不符合任何一項]

1.6 成語、諺語：成語依基本原則(1)一律合併，諺語則依輔助原則(6)將成分作切分。

例 1. 陰錯陽差、貌合神離、一不做二不休、一而再再而三

例 2. 話不投機 半句多、虎落平陽 被犬欺

2.各種詞型的分合層次

(1) 信級：凡是收錄在標準辭典中的詞一律斷開。

(2) 達級：能以構詞律組合出來的詞在達級合併。

(3) 雅級：無法完全收錄在辭典中的詞在雅級合併。

2.1 複合詞

2.1.1 定量式複合詞

(1) 數詞:合(達級)

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、七、六五八·四六、AB-8888、A110048787、7：20：30、第一、O二—七八八—三七九九—一五零一、7883799*、2/28—3/31、三十餘、一百多、二分之一 強、四十% 以上

註* 這是電話號碼。

(2) 表特定時間、地點之定量詞:合(達級)

例 1. 西元 一九九五年 三月 六日 二點 二十分、二十世紀

例 2. 八十學年、八十四學年度

例 3. 三年五班、五班

例 4. 七十巷 三十五弄 二號 之一 四樓 B1

例 5. 西經 一二〇度 四十六分

例 6. 四局 上、第四局 上 半

備考：時間詞中插要分(達級)

例：1995、96年；1995—1996年；本(二)月；七、八月；二至三月

(3) 普通定量詞：分

例：三 位、五十二 隻、三又二分之一 打、七十餘 位、七十 位 餘、六十多 國、三十來 歲、八 條 半、二十 個 左右

2.1.2 複合動詞

(1) 並列式：合成雙音節且不可前後互換才合(雅級)

例：醃泡黃瓜、發交相關單位、組建完畢、製播節目、依身材量製

(2) 偏正式：有衍生詞綴、接頭/接尾詞才合(達級、雅級)

備考 1.偏正式動詞之衍生前綴：可、好、互、相、自

2.偏正式動詞之接頭詞：加、改、重、增、轉、合、代、偷、抽、誤、速、趕、補、複、預、超、回、搶、借、試、大、小、共**、對、耐、續

註** 這裡的“共”表“共同”，非“一共”。

(3) 動賓式：辭典收詞才合（信級）

(4) 主謂式：辭典收詞才合（信級）

(5) 述補式

(a) 方向補語：合（雅級）

(b) 結果補語：補語是單音節：合（雅級）

(c) 結果補語：補語是雙音節：分

備考 1.方向補語指的是：上、下、過、起、開、回、進、出、上來、上去、下來、下去、起來、回來、回去、進來、進去、出來、出去、過來、過去

2.不論在語意上是結果補語或方向補語，動詞加方向補語一律合。

例：走回來、挽救回來。

3. 但是“起來”、“下去”作時態標記時不和前方動詞合。

例：保持 下去、尖叫 起來

4. “起來”作評價副詞時，和前方動詞合。

例：這件 衣服 看起來 不錯

(6) V-給、V-到、V-於、V-有、V-爲、V-成、V-作 另行規定

(a) V-給：合（雅級）；但動詞本身是動賓或述補結構：分

例：批發給、寫信 給、分紅 給、取出 給、退回去 給

(b) V-到：合（雅級）；但後接時間、補語和數量詞時：分

例：接觸到、聊 到 半夜、走 到 腿酸、加 到 兩百萬

(c) V-於：分；但動詞是附著詞、或合併後意義改變、或表示比較：合（信級）

例：生 於 台北、客於、有感於、大於、優於

(d) V-有：合（雅級）

(e) V-爲/成/作：合（雅級）

(f) V-不得/不了：合（雅級）

2.1.3 偏正式複合名詞

(1) 簡單式***

(1.1) 帶語法詞綴：分

名詞性語法後綴：者、們

(1.2) 帶衍生詞綴：合（達級）

(a) 衍生前綴：老、小、第、阿

(b) 衍生後綴：氏、某、度、性、家、長、師、員、兒、儿

(1.3) 2+1 音節：帶接尾詞就可合（雅級）

(1.4) 其他：符合下列標準才合

(a) 帶接頭詞（雅級）

例：副校長、準博士

(b) 含有附著成分（雅級）

例：奇案、勇將

(c) 語意無組合性（信級）

例：土包子、鐵公雞

(d) 專指（信級）

例：白菜、黑板

(e) 使用頻率高（信級）

例：牛肉麵

註*** 詞綴及接頭詞/接尾詞，請參照附錄 2。

(2) 複雜式

(2.1) 簡短、常見式：合（信級）

例：借書證、租車費

(2.2) 冗長、少見式：分

例：太空 計畫 室

2.1.4 複合介詞

(1) 辭典收有：合（信級）

例：改以

(2) 辭典沒收：分

例：親 至、親 與

2.1.5 複合副詞

(1) 辭典收有：合（信級）

例：正在

(2) 辭典沒收：分

例：並 非

2.1.6 名方式複合詞

(1) 辭典收有：合（信級）

例：空中、日前、月底、門口、街口

備考：如果名詞與其前面成分為一詞，則方位詞仍可能被切分開。

例：三月 底、大門 口、臨沂街 口

(2) 辭典沒收：分

例：手 內、腳 邊

2.2 專有名詞

(1) 單純詞：合（雅級）

例：胡適、桂林、布農、貝多芬、克寧、阿爾及利亞、宇宙光

(2) 專名+普名：

(a) 普名是接尾詞：合（雅級）

例：阿美族、光復橋、竹聯幫、大漢溪、桃園廠、王董

(b) 普名是自由詞素：分（雅級）

例：胡 先生、平漢 鐵路、二二八 事變、永新 加油站

(3) 縮寫：合（雅級）

例：勞基法、奧申委、文建會、臺三線、北二高、中常會

(4) 複雜詞：分

例：台北市 第一 信用 合作社、省 自來水 公司

(6) 詞組或句子：分

例：鯨魚 的 生 與 死、讓 我們 看 河 去

2.3 簡稱：合（雅級）

例：男單、女網、空姐、影視、化工、音像

2.4 合併詞

2.4.1 無中插

(1) 詞頭合併：合（雅級）

例：國內外、高中職

(2) 詞尾合併：合（雅級）

例：父母親、公私立

(3) 頭尾合併：合（雅級）

例：中山南北路

(4) 套裝合併

(a) 前面是專名：分

例：台北 市長、正義 里長、新竹 縣政府

(b) 前面是其他：合（雅級）

例：事務局長、體育司長

2.4.2 有中插：分

2.5 重疊詞

2.5.1 無中插：合（達級）

(1)動詞

(a) 嘗試貌：談談、想想、研究研究、說說 看

(b) 暫時貌：坐坐 就 走、擦擦 就 可

(c) 程度貌：胖胖 的、辛辛苦苦

(2) 名詞：車車、狗狗、小彬彬、痘痘

(3) 量詞：片片、一片片

(4) 擬聲詞：叮叮噹噹、乒乒乓乓

2.5.2 有中插：分

例：說 一 說、想 了 想

2.6 正反問句

(1) 動詞：分

(a) 完整形式：分

例：喜 歡 還 是 不 喜 歡、喜 歡 不 喜 歡

(b) 不完整形式：合（雅級）

例：喜 不 喜 歡

(2) 法相詞：合（達級）

例：能 不 能、可 不 可 以、可 以 不 可 以

2.7 否定式

(1) 不-動詞：分

(2) 不-法相詞：合（信級）

備考：有的法相詞也有動詞的用法，否定時若為動詞用法則予以切分。

例 1. 你 不可 走 / 你 非 走 不 可

例 2. 你 不應該 說 / 你 這 樣 說 不 應該

2.8 中插詞：分

(1) 動賓中插：分

例：洗 了 一 個 澡

(2) 述補中插：分

例：打 得 破、打 不 破

(3) 動賓、述補交互中插

(a) 相鄰者優先合併

例：洗 好 澡、剃 光 頭、吃 飽 飯、彎 下 腰 去

(b) 無相鄰者全分

例：幫得上忙、喘不過氣來

(4) 合併中插：分

例：初、高中；國內、外；中山南、北路

(5) 重疊中插：分

例：笑了笑、哭一哭

2.9 成語換字或固定可套換的詞組

(1) 辭典收有：合（信級）

(2) 辭典沒收：

(a) 組合成分的語法行為改變，或含附著語素：合（雅級）

(b) 不違背組合成分的詞類，且不含附著語素：分

例 1. 有氣無力（辭典收有）、有詢無質（含附著語素）、有橋無路

例 2. 連哄帶騙（辭典收有）、連誇帶褒（含附著語素）、連拉帶勸

例 3. 東張西望（辭典收有）、東蹶西振（含附著語素）、東填西寫

2.10 動後修飾語：合（信級）

(1) 辭典收有：合（信級）

例：僵持不下、井然有序

(2) 辭典沒收：分

例：對峙不下、排列有序

備考：動後修飾語有

不下、不力、不休、不羈、不成、不誤、不已 ...

有方、有加、有功、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 ...

附 錄 2

詞綴、接頭/接尾詞參考表

1. 詞綴

1.1 語法後綴：

- (1) 名詞的後綴：們
- (2) 動詞的後綴：了、著、過、看、看看
- (3) 關係子句標記：者、的/地

1.2 衍生詞綴

1.2.1 衍生前綴：

- (1) 名詞的前綴：老、小、第、阿
- (2) 動詞的前綴：可、好、相、互、自

1.2.2 衍生後綴：

- (1) 名詞的後綴：氏、某、度、性、家、長、師、員
- (2) 動詞性後綴：化
- (3) 多功能後綴：兒、儿

2. 接頭詞/接尾詞

2.1 接頭詞

- (1) 名詞的接頭詞：副、非、多、高、低、超、零、單、雙、前、準、總、主、代、巨、泛
- (2) 動詞的接頭詞：加、改、重、增、轉、搶、合、代、偷、誤、補、預、借、回、速、複、超、試、抽、趕、共、大、小、對、耐、續

2.2 接尾詞

- (1) 名詞性的接尾詞：

ㄅ 壩波簿盃碑杯輩包堡報班板版本幫榜棒別錶表濱病比幣壁埔部兵
 ㄆ 坡婆牌排派盤盆旁棚票篇片品瓶坪癖鋪舖
 ㄇ 碼膜饅模沫脈帽盟苗廟面皿迷木目
 ㄋ 法費販犯粉坊方房風峰服符
 ㄌ 地帶袋島道豆單彈黨檔調點殿店叮堆隊端段洞凍
 ㄊ 臺台態陶套頭攤灘壇潭譚堂帖廳亭艇題體團通筒桶徒圖土
 ㄋ 年能泥農女
 ㄌ 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率律
 ㄍ 閣糕膏稿鎬溝乾桿感缸岡崗港卦鍋國果糶鬼櫃關觀官管館罐功工
 宮公骨股鼓谷
 ㄎ 卡科客課考坑窟塊款狂孔褲庫

厂 盒河號函漢行會花話貨環荒鬚壺湖戶畫
 ㄩ 機肌夾呷假架價街結節界膠間艦鍵金江獎醬匠經精鏡級籍集計劑
 記祭季圈捲君軍郡局劇
 ㄨ 期橋球琴腔槍旗氣器區曲拳券權群裙
 ㄊ 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
 ㄓ 渣宅齋寨照罩州洲展站戰針陣鎮章症證証珠株桌傳庄裝莊狀主史
 冢種塚竹柱質值職誌製制
 ㄛ 癡痴差潮禪場廠城川串窗處池
 尸 生石舍社衫山扇身省商聲書刷數稅霜署術樹束市式室
 日 熱日
 ㄖ 仔澡作座罪圳宗族組子字
 ㄗ 側冊財操槽餐艙層村叢詞祠
 ㄘ 絲司色賽傘散酥所鬆素寺
 ㄗ 額
 ㄝ 癌
 ㄣ 案
 一 衣液頁業藥炎巖岩堰宴印營儀椅尾裔
 ㄨ 屋窩位味灣灣丸紋文王網舞物
 ㄩ 源園苑語嶼慾欲浴域樂

- (2) 非謂形容詞性接尾詞：用
- (3) 動詞性的接尾詞：給、有

附錄 3

操作型定義

信、達、雅三級是根據自動化困難度及實際使用情況而分層次處理，信級為文獻資料交換之基本層級，依據基本辭典中的詞項，將詞一一斷開，而沒有收錄在辭典的詞全部被斷成一個一個的字元，在此層次，除了可解決歧義，還符合基本資料交換的最基本形式。達級則為目前自然語言處理適用層次，可以由構詞律組合的在此層次處理、合詞。雅級為理想層級，所有應該合詞的全部在這個層次合詞。但是，在雅級，某些類型的詞仍可找出組合的規律，另一方面，雅級的自動化不易達成，因此達級和雅級之間可以處理的類型不硬性規定，改採浮動等級。採取浮動等級的好處有：(1)更可以符合不同需求，(2)以表列方式可以提供給使用者實際操作型的根據。以下為操作型定義，依照詞類、結構、次類、構型分類，每一類型並給實例，依「中文資訊處理分詞規範」登錄每一類型分或合的情況。

詞類	結構	次類	構型	實例	合/分
普通名詞	含附著語素		詞綴	<u>開創性</u>	合
			接頭/接尾詞	<u>集中點/次系統</u>	合
			後綴(接尾詞)+後綴(接尾詞)	<u>入山證費</u>	合
			前綴(接頭詞)+後綴(接尾詞)	<u>多面性</u>	合
			[簡稱]	<u>社青</u>	合
		複雜結構	偏正+詞尾	<u>緊急處分權</u>	分
			N+V(N為V的受詞)+接尾詞/詞綴	<u>生命孕育力</u>	分
			(-型最常見)	<u>帶狀新聞雜誌型</u>	分
			疾病名稱：(專門辭典收錄)	<u>性精神宜能症</u>	分
			疾病名稱：(專門辭典收錄)	<u>心臟病</u>	合
			不+V+接尾詞/詞綴	<u>不安全感</u>	分

			可 + V + 接尾詞/詞綴	可擴充性	合
			非 + V + 接尾詞/詞綴	非一面倒式	合
			無 + N + 接尾詞/詞綴	無症狀性	合
			最(更、特、極) + 高(低) + N	更高難度、極高點、最低處	合
	定 量 式		並列	十五二十	合
			排行(數字 + 姐、妹、哥、弟、伯、叔、姑、...)	四妹	合
	偏 正		1 + 1	土狗	** ¹
			1(前綴、接頭詞) + 2	早自修	合
			2 + 1(附著語素、後綴、接尾詞)	人工式、人事課	合
			2 + 1(自由語素)		分
			3 + 1(附著語素、後綴、接尾詞)	玉蜀黍油	合
			3 + 1(自由語素)		分
			2 + 2 [種類: 動物、植物、化學名稱、...]	大王椰子、灰面鷲鷹	**
	並 列		1 + 1(簡稱、...)(附著語素)	中西	合
			1 + 1(簡稱、...)(自由語素)	中美	分
			1 + 1 + 1(簡稱、...)	夏商周	分
			1 + 1 + 1 + 1	男女老幼	合
			2 + 2(偏正 + 偏正)	名山勝水	合
	合 併		詞尾合併	父母親、壯老年人	合
			詞尾合併(並列 + 接尾詞/後綴)	農林漁牧業	合

			套裝合併	體育司長、大板車伏	合 ²
				台北 市長	分
				勞工 行政 股 長	分
	簡 稱			社青	合
	重 疊		AA (辭典收錄)	人人	合
			AAB (AA 聲)	咕咕聲	合
			AAB (辭典收錄)	毛毛蟲	合
			AABB	山山水水、子 子孫孫	合
	固定可套換 之型式 ³		寸 N 寸 N	寸土寸心	合
			千 N 萬 N	千山萬水	合
			南 N 北 N	南腔北調	合
			一 N 一 N	一草一木	合
			一 N 半 N	一官半職	合
			前 N 後 N	前因後果	合
	含英文字		英文 + 接尾詞	BB 槍	分
			辭典已收錄的詞	卡拉 OK	合
	含標點符號	普通名詞	--、		分
定 詞	定 量 式		第 + 數字	第五	合
			數字 + 定詞	100 多 (vs. 九 成多)、20 幾、二十餘	合
			合併	三四百	合
	重 疊	量 詞	AA	包包	合
			一 AA	一包包	合
			二 A 二 A	一包一包	分
	含標點符號	數 詞	小數點符號、分數符 號、百分比...	92%、2/3、百 分之四.五、 三,八九二. 四...	合
地 方 名 詞	含附著語素	地 名 / 點	機關組織	行政院	合
			團體	經濟學界	合

			科系	化學系	合
			住址	中山北路	合
			地區(行政區域)	台灣省	合
			地名	水沙連	合
			地點	文武廟、水果行、呂宋島、加州灣、出貨區	合
			物品(頭、尾)	片尾、船尾、車頭、牙刷尾	合
			身體部位(頭、尾)	眼尾	合
		地名 / 點	[簡稱 + 後綴 / 接尾詞] 機關組織	亞銀	合
			[簡稱]團體	泳協、社女組	合
			[簡稱]學校	花師、雄女	合
			[簡稱]科系	人社院、動物所	合
			[簡稱]地區(行政區域)	川省	合
			[簡稱]地名	佛州、基市	合
	定 量 式			十八尖山	合
	並 列		1+1 (附著語素)	中西	合
			1 + 1 (自由語素)	英 法	分
	合 併		詞頭合併	高中職、國中小	合
			詞尾合併	上下游、南北韓	合
			頭尾合併	中山南北路	合
			套裝合併	濁水溪床、經濟學界	合
	重 疊	方 位	AAB	北北西	合
		地 點	AABB	內內外外	合
	含 英 文 字	地 點 / 名		<u>三</u> <u>E</u> 、 <u>BBS</u> <u>站</u>	分

	含標點符號	地 方	外國地名	L · A	分
時間名詞	含附著語素		詞綴	吉時	合
			1{今、昨、明、去、初、前、後、晚}+1	去歲、今晨	合
	簡 稱		年代	明、清	合
			季節	春、夏、秋、冬	合
			節日	秋節	合
	定 量 式		日期、年	1998	合
	並 列		1+1	明清	合
			1+1+1+1	春夏秋冬	合
	合 併		詞頭合併	星期一二	合
			詞尾合併	上下午	合
	含英文字			97 BC	分
	含標點符號			17:00、2/8、(八十三)、	合
				10.05.93、1-27-90、1/27/90、1863-1865年	分
專有名詞	詞 綴		人名(中國)	趙元任	合
			名字 + {姊、弟、兄、...}	登女姊	合
			姓 + {姊、弟、兄、...}	陳姐、王伯、魯君	合
			稱呼(姓/名+職稱)	陳老師、星雲法師	分
			中外古代帝后	齊宣王	合
		(非複雜結構)	書名、報章雜誌名、歌名、電影名、戲劇名	三國志、琵琶行	合
			星座名	人馬座	合
			組織名	社民黨、湖人隊	合

			族群名稱	雅美族	合
			宗教名稱	天主教	合
			船名	鐵達尼號	合
			車名	復興號	合
			品牌	勝利牌	合
			獎項	諾貝爾獎	合
	定 量 式		電話號碼、車號、帳 號、身份證字號	2308711、4560	合
			產品系列	點三八	合
			船名	永春三號	合
			車名	賓士四五	分
			機名	波音 747	分
	含 標 點 符 號		電話號碼	(02)27883799 , 2788-3799 , 02-2788-3799 , 800-1-2788-3 799	合
			車號	M8-3656	分
			地址、信箱	3~1	合
			章節	3.1, 3.2-1, 2· 3·5	合
			產品系列	A-95	分
			外國人名	班傑明·法蘭 克林	分
	重 疊		人名：(小)AA	(小)仙仙	合
			人名：ABB(姓+ AA)	謝娟娟	合
	複 雜 結 構	詞 組	書名、報章雜誌名、 歌名、戲劇名、電影 名	鯨魚 的 生 與 死	分
		句 子	書名、歌名、戲劇 名、電影名	讓 我 們 看 雲 去	分
動 詞	詞 綴			好吃、可賀、 癌化	合

	接頭 / 接尾 詞			大敗、共樂、 耐穿、留有	合
	偏 正		方位詞(上、下、前、 後、左、右、東、西...) + V	南徙、東移	合
			接頭詞、(久、分、 互、不、雙)+單音 節 V	續跌、分贈	合
			數字 + V	百忙	合
	並 列		1 + 1	組建、騎坐	合
	主 謂		1 (N) + 1 (V)	獸交、口臭	合
	動 賓		1+1(附著語素)	聽訓	合
			1+1(自由語素)	買書	分
	述 補	動作動詞	V + 方向補語 (上、 下、起來、下去...)	走上、躺下、 站起來	合
			V + 單音節 V	調鬆	合
			V + 雙音節 V	看清楚、填 寫完整	分
		作格動詞	V + (紅、壞、僵、 呆、死、垮)	唱紅、哭紅	合
	合 併		詞尾合併	上下班、接送 機	合
	重 疊 詞	動作動詞	AA	打打	合
			AAA	嘻嘻嘻	分
			AAAA	對對對對	分
			AABB	縫縫補補	合
			ABAB	分析分析	合
			ABB、V 光光	繞圈圈、吃光 光	合
			ABB (辭典收錄)	笑嘻嘻	合
		(動 賓)	AAB	點點頭	合
			AABC	獻獻殷勤	合
		狀態動詞	AA	冷冷	合
			AAA	羞羞羞	分
			AAAA	慢慢慢慢	分

			AAB (好好 V、小小 X、其他)	惶惶然、好好玩、小小聲、偷偷懶	合
			AAB (辭典收錄：-然、無相對應的 AB 詞項)	暈暈然、強強滾	合
			AABB	彎彎曲曲	合
			ABAB	悠哉悠哉	合
			ABB	濕淋淋	合
			A 裡/里 AB (辭典收錄)	怪裡怪氣、糊里糊塗	合
	正 反 問 句		A 不 A	辦 丕 辦	分
			A 不 AB	講不講理	合
			AB 不 AB	講理 不 講理	分
	固 定 可 套 換 之 型 式		一 V 一 V	一來一往	合
			一 X 二 X	一清二楚	合
			一 X 半 X	一知半解	合
			入 N 入 N	入情入理	合
			大 V 大 V	大起大落	合
			大 V 小 V	大呼小叫	合
			不 X 不 X	不聞不問、不男不女	合
			一 V 而 V	一飲而盡	合
			不 V 而 V	不歡而散	合
			A 來 A/B 去	瞧來瞧去、丟來拋去	合
			亦 V/N 亦 V/N	亦步亦趨、亦歌亦泣	合
			上 V 下 V	上沖下洗	合
			V 上 V 下	跑上跑下	合
			東 V(單音節) 西 V(單音節)	東奔西跑 東逛逛西晃晃	合
			V 東 V 西	擦東擦西	合

			V 長 V 短	說長道短	合
			南 V 北 V	南來北往	合
			沒 X(單音節) 沒 X (單音節)	沒心沒肺	合
			七 X 八 Y	七老八十、七手八腳	合
			千 V 萬 V	千叮萬囑	合
			隨 V 隨 V	隨到隨辦	合
			前 V 後 V	前呼後擁	合
			V 前 V 後	思前想後	合
			V 長 V 短	截長補短	合
			左 V 右 V ⁴	左思右想、左等右等	合
			半 X 半 X	半工半讀	合
			半 X 不 X	半大不小	合
			似 X 非/未 X	似懂非懂、似明未明	合
			似 X 似/若 Y	似真似幻、似幻若真	合
			如 X 似 Y	如饑似渴	合
			如 V 如 V ⁵	如癡如醉	合
			V 而不 V	視而不見	合
	中插結構	動賓中插	中插「了、著、過、」及定量式	洗了二個澡	分
			沒有相對應的動賓動詞(辭典收錄)	行大運、鬆口氣	合
		述補中插	V(述) 不/得 R(補)	打得破、打不破	分
			沒有相對應的述補動詞(辭典收錄)	管不著、氣不過	合
		動賓、動補交互中插		彎下腰去、喘不過氣來	**6
		合併中插		上、下課	分
		重疊中插	A 了 A、A 二 A	看了看、縫一縫	分

	V - X	VD	V-給	敬獻給	**7
			V-到	接觸到	**8
		VJ, VCL	V-於	吝於、有感於、大於	**9
		VG	V-爲/成/作	轉化爲、寫成、當作	合
			V-不了	控制不了、逃不了	合
			V-得了	管得了、解決得了	合
		VJ	V-有	存有	合
		VB, VH	V-開來	分割開來	
			V/D(雙音節)-不已	驚訝不已、憤憤不已	合
			V/D-不平	崎嶇不平、憤憤不平	合
			V-不休	爭論不休	合
			V-兮兮	緊張兮兮	合
			V-不絕	洶洶不絕	合
			V-不下	僵持不下	合
			V-不得	歸不得、碰觸不得	合
	含英文字	動賓	英文+中文	<u>K</u> 他	分
		偏正	英文+詞綴	amplify 到	分
			辭典已收錄的詞	K 書	合
	成語		(辭典收錄)	三陽開泰	合
非謂形容詞	接尾詞		(-式、-用、-等、-流、-製、...)	日式、農用、銅製、	合
	主謂		1+1 (-營)	市營	合
	重疊		AABB	真真正正	合
	合併		詞尾合併	公私立	合
	偏正		-屬 (Nc /Na 屬 X : X 屬於 N) (1+1)	市屬	合
			-有 (Nc 有 X : X 歸 Nc 所有) (1+1)	鎮有	合

副詞	重疊	擬聲詞	AA	呵呵	合
			AAA	咯咯咯	分
			AAAA	噹噹噹噹	分
			AABB	滴滴答答	合
			ABAB	戈登戈登	合
			AAB	叮叮噹	合
			ABB	嘩啦啦	合
		時間副詞	AA	每每	合
		方式副詞	AABB	踉踉蹌蹌	合
			ABB	惡狠狠	合
		其他	ABAB	千萬千萬	分
	正反問句		A不A	會不會	合
			A不AB	應不應該	合
			AB不AB	應該不應該	合
	偏正法相詞		不+D	不可	合

註⁽¹⁾ ** 表分合需根據其為附著語素或自由語素來決定。

註⁽²⁾ 套裝合併是前一詞的末尾與後一詞的開頭相同，可不再重複出現。形成套裝合併的詞多半包含表示行政區域、單位或表示階級的接尾詞。套裝合併可依下面原則切分：

- (1) 不含行政區域的單純套裝合併詞予以合併。例如：建設局長
- (2) 含地名或行政區域的單純套裝合併詞予以合併。例如：台北 市長
- (3) 複雜合併詞套裝依成分詞切分。例如：勞工 行政 股長

註⁽³⁾ 固定可套換之用語多半為四字詞。

註⁽⁴⁾ 左鄰右舍為名詞，收錄在辭典中。

註⁽⁵⁾ 「如痴如醉」合詞，其中「痴、醉」為動詞。但是「如詩如畫」則切分成「如 詩 如 畫」，因為「如 N 如 N」，而且可以由二個以上的成分組合，當成並列形式的組合，意義不改變，如 父 如 兄 如 母 如 姊。而如 V 如 V 的類型，則固定由兩個成分組合。

註⁽⁶⁾ 動賓、動補交互中插的例子也很常見，目前規定相鄰者優先合併，而無相鄰者分，例子：彎下 腰 去（述補相鄰）、睡 丌 著 覺（不相鄰）。

註⁽⁷⁾ V-給：合，但動詞本身為動賓或述補結構時，分。例如：

泡茶 給（動賓結構+給）、提供出來 給（述補結構+給）

註⁽⁸⁾ V-到：合，但後接時間、補語和數量詞時，分。例如：

聊 到 凌晨 兩點（時間）、賺 到 三百多萬（數量詞）、鞠躬 到 九十 度（數量詞）、
走 到 腿酸（補語）

註⁽⁹⁾ V-於：分，但動詞本身為附著詞時，或合併後意義改變、表示比較時，則合詞。例如：

生 於 台北、客於（「客」為附著詞）、有感於（合併後意義改變）、大於（表示比較）。
