

Several Discussions on Chinese Letter-word phrases*

Zheng Zezhi¹ Zhang Pu²

1. Xiamen University, Xiamen, Fjian, China
2. Beijing Language and Culture University, Beijing, China
zezhi@126.com; zhangpu@blcu.edu.cn

Abstract

Based on the investigation of the usage of letter-word phrases in the “People’s Daily”(year 2002), several discussions, including classifying ELWPs occurring in the corpus, defining the ELWPs, analyzing and categorizing mono-alphabetic ELWPs and digital ELWPs, exploring ELWP parallel structures and so on, on letter-word phrases are presented in the paper. Wish to bestead the Chinese information processing and letter-word phrases standardization.

Keywords

ELWP, Letter-word phrase, mono-alphabetic ELWP, digital ELWP, parallel structure

1 Introduction

Since 1994 letter-word phrases have been studied, Yongquan Liu published the first paper on lettered-words “Discussing lettered-words”(Yongquan Liu, 1994.10), then along with lettered-word phrases come forth, more and more people pay attention to the new language phenomena, and the study goes widely from the form to the pronounce and semantic researches of letter-word phrases. These days, Southeast Asia Chinese districts and China are the main areas using letter-word phrases. The leading research fields about letter-word phrases are how to standardize lettered-words, how to select them for dictionaries, and how to read them.

Although there are many people studying letter-word phrases, almost all of their studies are illustrational or of qualitative analysis, and the researches in quantitative analysis + qualitative analysis are rare, much less reports about the usage of letter-word phrases in large scale real texts. By now we haven’t known clearly the sematic classes,

* This is a revised version of the paper presented at the 6th Chinese Lexical Semantics International Congress of Linguists, held at Xiamen, China in April 2005. I am grateful for the imbursement from the project (serial number: 04L2004-01-01-03) of the Research Center of National Language Monitoring Agency (press & books), which funded the research reported in this study. The authors are responsible for any remaining errors and inadequacies in the paper.

pragmatic state and the using range of letter-word phrases. We have investigated the usage of foreign letters in “People’s Daily”(year 2002)(Runzhi Guan, et al, 2005; Zezhi Zheng, et al, 2005), and based on the investigation this paper will give a external describing of the usage of foreign letters in Chinese texts, and then discuss mono-alphabetic ELWPs, digital ELWPs, ELWP parallel structures, and we wish the study to be beneficial to the Chinese information and standardizing letter-word phrases.

2 The definition of ELWPs

For lettered-words, Yongquan Liu ‘s definition was (Yongquan-Liu, 2002): “The words which are composed of Roman alphabets (including Chinese pinyin alphabets) or Greek alphabets or a composite of Roman alphabets and symbols and figures and Chinese characters or a composite of Greek alphabets and symbols and figures and Chinese characters.” This definition basically includes main properties of lettered-words. But it lacks in practicability for auto-extracting and recognizing lettered-words.

The reasons why we use the term “Letter-word phrases”, not “lettered-words” are: 1. The letter strings in lettered-words are almost the breviaries of foreign words; 2. Most time letter strings require joining with Chinese characters to denote a concept, for instance, to express a proper noun or a term; 3. The dividing line between Chinese words and phrases is vague.

As a matter of fact, neither lettered-words nor Letter-words phrases can include all usage of foreign letters in Chinese texts, such as 勋(音:Guī), v+m, .dbx, .dll, .doc, .eml, (C), C, [M], c: \io.sys, (c), c: \kkk, c: \windows, c: \windows\win.com, C: \Winnt\sys—tem32 etc. They aren’t a lettered-word or a Letter-word phrase. So we have to know all usages of foreign letters to make clear what are Letter-word phrases. Herein we present the definition—engineering definition of lettered words or phrases, for short, ELWP. The definition of ELWPs, it mainly refers to these character strings, which appear in Chinese texts, and are consist of word-symbols and mark-symbols, or word-symbols and mark-symbols and Chinese characters. The character strings have definite sense and syntax function (such as 卡拉OK, CD盘, V I S A 卡, H S K, 3 D动画 ISO9000 认证, IEC 标准, etc.). As a word or a phrase, it is self-contained, and in texts the letter sequence can’t be changed, its composition can’t be inserted or deleted.

In the definition, we use two terms, as follows:

- Word-symbol: It refers to “the minimum unit of pinyin characters or phonetic notation symbols”(GB/t12200.2-94, the part II of “Chinese information processing glossary”: Chinese and Chinese characters), it includes Roman alphabets, Greek alphabets, Cyrillic alphabets, Nipponese katakana and hiragana, etc.
- Mark-symbol: It refers to punctuation marks, money-symbols(\$, ¢, £, ¤, ¥, €, \$, \$, ¢, £, ¥, etc.), measure-symbols(°C, °F, №, ™, %, mil, mg, log, kg, ln, mm, KM, cm, cc, km, m², etc.), numeral symbols (Arabia numeral and Roman numeral), calculating-symbols(+,-,/,etc.) and other symbols(№, ™, &, *, #, ©, ®, etc.).

According to the definition of ELWPs, an ELWP can be an Internet address, an Email-address, a computer file name or a computer file address, a computer virus name, a formula, which contains Word-symbols, a figure + measure unit, foreign words or phrases, the Chinese pinyin, all kinds proper nouns (criterion names, agreement names, commodity names, brand names, company names, coding names, etc.), lettered words, etc.

In order to auto-recognize ELWPs, we have also presented a formalization definition of ELWPs (Zezhi Zheng, Pu Zhang, 2005) Thereinafter, we use ELWPs to denote the engineering definition of lettered words or phrases, and use lettered-word phrases to denote these words or phrases, such as “DVD 机, CT, etc.”.

Tab1 Some examples of ELWPs

ELWP	-ELWP
窄频 CDMA 日本电话电信公司 DoCoMo “A. O. C” 字样 格 C 盘 (Harm. FormatC. bwp20) 天然维生素 E http: //2002. people. com. cn “教育考试服务中心” (EducationalTestingService, ETS) c: \bbb 著名品牌 “EPSON 爱普生”	丰富的 维生素 E 、 美国教育考试服务处 (ETS) 举办的、 给 “E 学生”、 著名品牌 “EPSON 爱普生” 的商标 销售假冒 EPSON (爱普生) 墨盒 唱 “卡拉 OK” 中国横向长江经济带 “ T ” 型结构交汇点

3 the usage state of foreign letters in Chinese texts

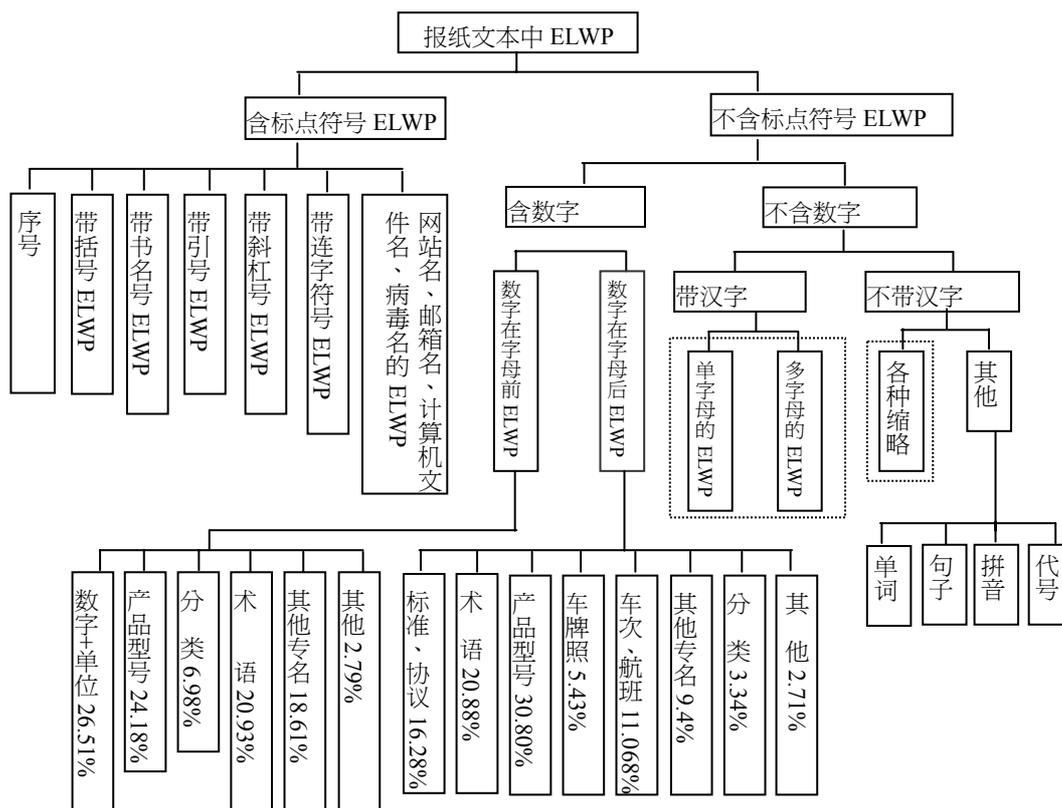


Fig.1 The classified ELWP in newspaper texts

According to the definition of ELWPs, we examined the usage of foreign letters in Chinese texts, and then classified the usage state (shown in fig.1). Heretofore Yongquan Liu(2002) mentioned 5 classifying methods. Our divisional way is based on the usage state of foreign letters, and from the angle of engineering view. This divisional way is valuable to the Chinese information processing and the criterion of letter-word phrases.

In Fig.1, the part containing statistical values is the part that we'll analyze. To the ELWPs with punctuations can be saw in the paper by Runzhi Guan et al(2005).

he ELWPs in Fig.1's big weak frame—the ELWPs with Chinese characters and without figures and punctuations are letter-word phrases, which are granted by many academicians. The ELWPs in Fig.1's small weak frame are regarded as breviaries of foreign words, and not lettered-words (Mingyang Hu, 2002). We consider these things resemble to the ELWPs in Fig.1's big weak frame, and should be regarded as letter-word phrases or lettered-words.

4 The mono-alphabetic ELWP

The mono-alphabetic ELWPs are the ELWPs which just contain only one word-symbol, such as “e 龙公司, G 网, K 粉, Q 号, C 盘, D 大调小提琴协奏曲, A 师, e 交通, etc.”.

The mono-alphabetic ELWPs, in letter-word phrases, has its particularities, A, B, C, D, etc. these letters can be used as serial numbers, class code names and may form a letter-word phrases when they combine with Chinese characters, so we examine them solely, and wish to find their sematic and pragmatic usages. The letter roles in mono-alphabetic ELWPs are shown in Fig.2.

There are 358 different mono-alphabetic ELWPs (used 1545 times, in 1138 texts), in the “People's Daily”(year 2002).

According to tab.2, a single letter is used to indicate a sequence number, a class code, and a team number in ELWPs, because in a way all these three usages of a single letter have taxis function, sometimes we can't class them clearly, certainly, some of them can be distinguished entirely, such as:

- As a team number: C 组,世界杯 D 组,世界杯 G 组,E 组,F 组,G 组,H 组,N 组, etc.
- As a class code: F 字签证 (访问类签证), A 型乙肝,摩托罗拉 V 字头 (摩托罗拉 V 字头是时尚类手机), etc.
- As a sequence number: A 级,A 级标准,A 级影片,D 级危房, etc.
- What are the different between “D 类危房和 D 级危房”? the letters in “维生素 A、维生素 B、维生素 C” are sequence numbers, also are class codes, and same as in “印楝素 A、印楝素 B、印楝素 D”.

A letter used as a symbol, such as: C 馆, C 国, C 盘, C 票, C 区, A 店, A 国, A 里, 阿 Q, 阿 Q 精神, B 立柱, etc.

Borrowing form ELWPs, using a letter shape to form a letter-word phrase, such as “H 型钢, O 型着陆, S 弯度, S 形, S 形流水, S 形平行状, S 形舞, T 型台, T 恤, T 恤衫, T 字台, etc.” These letter-word phrases may come from Chinese, such as “T 型台”, may come from foreign language, for instance “T 恤衫”.

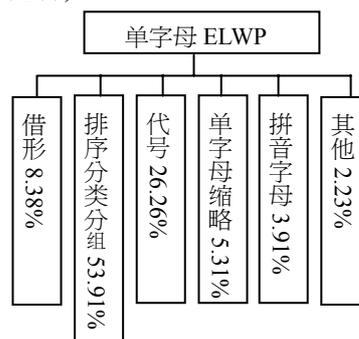


Fig.2 The letter role in the mono-alphabetic ELWP

Tab.2 the state of mono-alphabetic ELWPs

分类 数量	借形	排序分类 分组	代 号	单字母 缩略	拼音 字母	其 他		
						并列结构	单位	文本错 误
种数	30	177	111	19	14	4	1	2
百分比	8.38%	53.91%	26.34%	5.31%	3.91%	1.11%	0.28%	0.56%

A foreign word abbreviated to one letter, and then come into being an ELWP, such as “C 网（窄带 CDMA 的简称）, e 城便利站, e 城便利站终端机, E 电视台, E 化, e 交通, e 教育, 四通四 S, 窄带 C 网, etc.” The number of such ELWPs is not so much, but they are common letter-word phrases.

Pinyin letters, for instance “学习 b, p, m, f, d, t, n, l 八个声母, a, o, e, i, u, ü 六个韵母.”. Why a pinyin letter can't be a lettered-word or a letter-word phrase is that it has no meaning, just used as phonetic notation, but they can be used in texts and are our examining objects.

Parallel structures, refer to these mono-alphabetic ELWPs which are used in parallel structures express in texts, such as “A、B、C3 组, A、B、O 血型, A、B 角制, A、B、H 流通股, 准驾车型为 A、B、N、P, etc.”. When be off their sentences, the letters will lose their meanings, but can be normally used in sentences. They should be lettered-words or letter-word phrase.

Unit name, the letters in mono-alphabetic ELWPs can be used as a unit name, for instance “K” in the sentence “理论上也有 100 多 K, ……”.

Chinese pinyin letters are not lettered-words or letter-word phrases, but when they are used as a abbreviation, such as “HSK (the test of Chinese standard)”, we consider them to be lettered-words or letter-word phrase, because from the word form to usage, their functions are similar to the abbreviations from foreign words.

5 The digital ELWP

Arabia numbers have being an impartible part in Chinese depiction, but none has examined the usages of the blendwords of letters and numbers or letters and numbers and Chinese characters in Chinese texts. In this section we'll realistically describe and discuss the usages, based on the investigation of digital ELWPs in the “People's Daily”(year 2002), and wish to know something of them.

In this paper the digital ELWP refer to these ELWPs, with numbers and without punctuations, such as “Win2000, 65MW, 2B 铅笔, 丰田 8A, 运 8F400 飞机, system32, TOP500, CDMA2000 标准, 摩托罗拉 V70 手机……”.

There are 691 different digital ELWPs used in the “People's Daily”(year 2002), they occupy 10% in total ELWPs, and they could be divided into two groups, one group in which the numbers are prior to letters and another in which the numbers are posterior to letters. So we can examine the different functions when numbers are at different position in ELWPs.

5.1 The ELWPs in which the numbers prior to letters

There are 212 different ELWPs (used 293 times, in 251 texts), in which numbers are prior to letters, in these ELWPs, the strings of numbers +letters has 3 usages:

- 1) As code numbers of product names or product models, such as “丰田 8A, 运 8F400 飞机, 5ELX, 5V 电喷发动机, 600MW 机组, 603G3503 花岗石.....”.
- 2) As breviaries of speciality terms, such as “3G 标准, 3G 网络, 2MDDN 专线, Oct4 基因, 4S 网点.....”.
- 3) As strings of figure+unit, such as “15ml, 167×600KW, 16V, 1800MHz, 1G, 272GB.....”.

The state of these ELWPs shows in Tab.3

Tab.3 the state of these ELWPs in which numbers prior to letters

	数字+单位	产品型号	分组分类	术 语	其他专名	其他
种数	56	51	14	45	40	6
百分比	26.51%	24.15%	6.98%	20.95%	18.61%	2.79%

The strings of figure+unit, such as “100Hz, 100M3, 100mg, 10KV, 10m, 10mg, 10MW, 11M, 1250GB, 140TB, 150CC 组, 150mg.....”.

77.78% term ELWPs are shortened form, such as “国际 3G, 3G 技术, 3G 基础设施服务, 3G 时代, 3G 市场, 3G 网络, 同步 3G 移动通讯, 3G 移动电话, 1X, 1X 系统.....”.

66.67% product model ELWPs use figure to represent quantity and the strings of figure+unit to represent product models, such as “宝马 745h, 日本东芝 800mA 电视显像透视镜, 火龙牌 90W 电热毯, 600MW 机组.....”, and the others use figures to represent code numbers or some measurement, such as “603G3503 花岗石, 丰田 8A, 运 8F400 飞机, 3PE 防腐钢管, 2B 铅.....”.

Other names, most of them are company names, conference names, technique names, agreement names and criterion name, such as “3Com 公司, 3M 公司, 3G 标准, 3G 无线传输标准, 3G 牌照, 欧洲 3G 牌照, 开放式 3G 平台, 3G 用户, 3GPP 标准会议, 3GPP 工作组会议, 3GPP2, 3GPP2 会议, 3GPP2 全会.....”.

Only 3B(a credit grade of fund) is not grouping in grouping and classifying ELWPs, and the others all use the strings of figure+unit to denote groups or classes, such as “125CCA 组, 125CCB 组, 125CC 组, 1600CC 组, 250CC 组, 3B, 国产 125CC 组, 国产 150CC 组, 国家 3A 级旅游度假区, 专业 125CCA 组.....”.

Other ELWPs include expressions(such as, 3+X), common word phrases(such as, 附件 2B) and disconnecting of ELWP parallel structures (as 鑫诺卫星 2A, 3A 转发器).

5.2 The ELWPs in which numbers posterior to letters

There are 479 different ELWPs (used 713 times, in 572 texts), in which numbers are posterior to letters, in the “People’s Daily”(year 2002). In these ELWPs, the strings of letters+figures are used as code numbers or symbols, and no letters is used as measure units, just a few are used as breviaries. The state of these ELWPs shows in Tab.4.

Tab.4 the state of these ELWPs in which numbers posterior to letters

	标准 协议	产品 型号	术语	牌照	车次、 航班	其他 专名	分组 分类	其他
种数	78	148	100	26	53	46	16	12
百分比	16. 28%	30. 80%	20. 88%	5. 43%	11. 06%	9. 6%	3. 34%	2. 69%

Standard and agreement names, such as “ISO9002 质量体系认证, CDMA2000 标准, CMM2 级资格, EGQS2002 监督认定指标, SO205 大豆合约, 标准 ISO9002 质量体系, 国际 ISO9002 质量体系, 美国 QS9000.....”

Product models, such as “奥迪 A6, 摩托罗拉 A6188, 帕萨特 B5, 裕晟 BLDG201 号, 别克 GL10 商务公务旅行车, 别克 GL8, GL8, 上海通用别克 GL8, GX470 多功能车, H986 设备, HBT40 混凝土输送泵, 飞利浦 HQ6885, HQ6885 电动剃须刀, 震撼 JS100, ZK6118H 系列客车, 大型空客 A300 飞机, A318 型.....”

Terms, such as “黄曲霉毒素 B1, 基因 P16, 金昌 EX6000 印花电脑设计分色系统, 巴黎股市 CAC40 指数, 马 c6, 数据库管理系统 DM3, 数字程控交换机 ZX500, 通用中央处理器 CPU 龙芯 1 号, 维生素 B1, 维生素 D3, 狭义货币 M1, 现金 M0, 血清甲状腺激素 FT3.....”

License plates, such as “川 AK3925, 浙 AT6142, 川 AV6457, 冀 F52366, 晋 M16672

Scheduled train and flight numbers, such as “T21 次, T55 次列车, K105 次, K157 次, L1 次列车, CA1225 次, CA129 航班, CA129 班机, 国航 CA129 航班.....”

Other names, 86.67% of them are code numbers, such as “中国海关 H986 工程, IBM4300 系列电子计算机, INTELSAT904 卫星, MD82 飞机, PA18 网站, 男子游泳 S12 级, S212 省道, 北环 N04 标段, SQR7160 车身, 乒乓球男子 TT9 级团体比赛, 女子乒乓球 TT9 级项目, ZXF118 固定电话网短消息中心.....”, the others use figures to denote amount, such as “F4, F4 组合, G7 成员国, G8 首脑会议.....”.

In the grouping ELWPs, the grouping gist is code numbers, such as “O I 类, T44 级, M2 组, N2 组.....”.

The others, 12 different ELWPs, among them, there are 11 pieces of pure code, such as “M01765, ROOM18, QQ1506414”, 1 piece text error (WT0 成员, the third character should be “O”, not “0”).

5.3 Our humble opinion on digital ELWPs

The letters in digital ELWPs, in which numbers are prior to letters, are used as measure units, class code, group code, grade codes, the figures are used as quantity, sequence number, etc., such as “600MW 机组, ‘3G’(for short, the third generation of motion communication, ‘3S’(for short, Geographical Information System, Remote Sensing, Global Positioning System), and the blends of figures and letters are used as strings of figure+units, product models or specific terms, etc.

The letters in digital ELWPs, in which numbers are posterior to letters, generally can't be used as measure units, the blends of figures and letters are commonly used as all

kinds codes, for example, standard and agreement names, product models, Scheduled train and flight numbers or technology terms.

Sum up the two usages of digital ELWPs, we can come to the conclude that the blends of figures and letters or figures and letters and Chinese characters are mostly used to express names or terms with code or code number in every walk of life.

About these ELWPs our view is as follows:

- a) Names and terms have their references and meaning, such as “MR479Q 发动机, A380 飞机, 办公软件 WPS2002 版, ISO14000 认证, 法兰克福 DAX30 指数, 黄曲霉毒素 B1, etc.”. They are international or institutional universal expressions, and have been used in Chinese for a long time, and these expressions are simple and clear, should be accept as letter-word phrases.
- b) The strings of figure+letter express amount, the letters are used as measure units. The foreign measure units after Chinese transformation have been a part of Chinese glossary, such as “千瓦, 兆, 欧姆,etc.” (see, “Modern Chinese Dictionary”, 1985). The letter-units Hz, GB, etc. are of internationalism and convenient, using in texts is more economical than Chinese transformation units, so they can be regard as letter-word phrases, and these ELWPs are controllable by rules.
- c) The blends of figures and letters used as Class codes, grouping numbers and common code are easy to express hiberarchy structures. They are natural uses of foreign letters, being regarded as letter-word phrases is acceptable.

Names and terms take up 78.83% in the Digital ELWPs, this figure shows that a great lot of letter-word phrases are names and terms, and also shows that products and techniques are the main source of letter-word phrases.

6 The ELWP parallel structures

Yunfang Wu (2003) studied Chinese parallel structures. For substantive parallel structures, she said: The rightest word in all parallel structures is the tolerant center phrase, while parallel elements are pure words or phrases, themselves are center words. If ELWP parallel structures have center words, the center words commonly should combine with letter-word phrases sequences.

In order to recognize ELWPs, we must affront ELWP parallel structures, therefore in this section we will discuss the problem based on the examining of ELWP parallel structures in the “People’s Daily”(year 2002).

The state of ELWP parallel structures in the “People’s Daily”(year 2002) is shown as Tab.5.

Tab.5 The state of ELWP parallel structures in the “People’s Daily”(year 2002)

	总出现次数	出现种数	光杆并列	只带后中心语	只带前限定	前限定后中心语
出现次数	142	127	75	40	4	8
占并列结构百分比	100%	100%	59.05%	31.5%	3.15%	6.3%
占总 ELWP 的百分比	1.21%	3.12%	1.84%	0.98%	0.098%	0.197%

Examples of ELWP parallel structures:

1) Pure ELWP parallel structures

Route.exe、Ftp.exe、Recover.exe、ScanReg.exe、D3d.dll、Sysscan.exe
WCDMA、CDMA2000
B级（帕萨特）、C级（奥迪A6）
DVD—Video、DVD—R、VCD、CD、CD—R/RW

2) Comma ELWP parallel structures

WTO, APEC, CBD, ATM, NMD, CEO, DNA
DY—2000, DY—2001, DY—2006 鹰系列可视电话
163.com, 163.net, 263.net, sina.com, china.com

3) Only with pre-determiner

海基导弹 SS—N—8、SS—N—18
俄罗斯陆基导弹 SS—18、SS—19

4) With post-center phrase

Windows9X、WindowsNT、Windows2000、WindowsXP 等操作系统
A、B、O 血型
A、B、H 流通股
T63、T106 数值预报模式
便携式 CD、MD 唱机
鑫诺卫星 2A、3A 转发器

5) With Chinese parallel structures symbol words

PCVD、MCVD、OVD 和 VAD
U—17 和 U—19
西耶那 EDX、EL 和 ELX 共 3 款车型
 α 和 β 受体
P O S 或 A T M 机
全球两大信用卡组织 V I S A 和 M A S T E R
“F”或“L”签证

If we save ELWP parallel structures as ELWP, we'll gain ELWP parallel structures, not

letter-word phrases, and if we disconnect the parallel structures, some letter-word phrases will be damaged in meanings or the expression integrality. Such as:

- 1) “T63、T106 数值预报模式” disconnected to: “T63”, “T106 数值预报模式”
- 2) “海基导弹 SS—N—8、SS—N—18” disconnected into “海基导弹 SS—N—8”, “SS—N—18”
- 3) “便携式 CD、MD 唱机” disconnected into “便携式 CD”, “MD 唱机”
- 4) “鑫诺卫星 2A、3A 转发器” disconnected into “鑫诺卫星 2A” “3A 转发器”
- 5) “ α 和 β 受体” disconnected into “ α ” “ β 受体”

For examples 1), 2), their meanings or the expression integrality are damaged, but every word is correct in meanings. For example 3) there are no part-of relation between “CD” and its determiner “便携式”, “便携式” just modifying “CD”, so every part is correct. Example 4) “鑫诺卫星 2A” is wrong in meaning, the reason is that the relation between its center word and determiner is a part-of relation. When we disconnect an ELWP parallel structure, and there are part-of relations in the structures, this operation may bring some errors.

By our analysis, for ELWP parallel structures, our processing policy is disconnecting the structures, in the worst, it may bring less than 0.3% absolute error rate.

As time lapses, Only “、” ELWP parallel structures may be saved using its distortion or its parallel structure, such as “A、B、O 型血”.

7 our view on letter-word-phrases standardization

Based on our examining, we find that the two cases should be regulated:

The first thing is abusing ELWPs, which refer to larding foreign words in Chinese expression, such as “倘若有一天, 连‘我爱你’都缩写成了‘ILU’, 我们除了‘I (我) 服了 U (你)’之外还能说什么呢?”. Such sentences should completely use Chinese words or phrases.

The second thing is spelling errors or different writing forms, including ELWP unmerited writing forms, spelling errors, punctuation errors, even misconstructions, etc. As a result, these errors bring us hetero-graphy ELWPs, homo-graphy ELWPs, which is a noticeable thing.

Homo-graphy ELWP refer to theses ELWPs, which have same forms and different meanings. Such as “NLP (Neuro Linguistic Programming), NLP (Natural Language Processing)”. These ELWPs can be used in text with bracket annotation, which can't bring any mistake to the ELWPs auto-extracting.

Hetero-graphy ELWP refer to theses ELWPs, which have different forms and same meanings. Such thing can be divided into 4 classes, as follows:

- a) Different origins and different users bring different forms, such as “E—mail, E-mail, e-mail, Email and mail”, “GPS 全球定位仪, GPS 全球卫星定位系统, GPS 卫星导航定位系统, GPS 卫星导航”, “DVD 和 DVD 机”, “mini DVD and 迷你 DVD”, etc.

- b) Different capital letters, such as “《TRIPS 协定》 and 《Trips 协议》”.
- c) Different punctuations, such as “《马拉喀什建立世界贸易组织协定》(《WTO 协定》) and 《马拉喀什建立世界贸易组织协定》 (“《WTO 协定》”)”, “Win-dowsNT, WindowsNT”, “WPSof-fice, WPSoffice”.
- d) Different space numbers, such as “《WTO 法律知识 100 问》 and 《WTO 法律知识 100 问》”.

Such ELWPs need to be standardized. To solve the Hetero-graphy problem, for d), we can use computer to finish; for a), b), c), according to the constructional roles of Chinese word phrases and the source form in foreign texts, and then set down a criterion to standardize letter-word phrases' forms and usages.

In addition, letter-word phrases have a special case, such as “CD(光碟), Cd(元素‘镉’的符号),cd(光学强度单位名称‘坎(德拉)’的符号)” (Yongquan Liu, 2002).

8 Conclusions

Our investigation is based on Chinese texts, thus don't touch upon the speech of letter-word phrases, this does not show that the speech is unimportant, just means that we haven't do it. The contrastive researches on letter-word phrases, which focus on different medium, different districts and different ages and so on, are our farther tasks.

References

- Runzhi Guan, Jianguo Yang, 2005, Interpunction Issues in Chinese Lettered Chunks, *Applied Linguistics*, no. 1. pp. 82-87.
- National Technology Superintendence Bureau, 1994, GB/t12200.2-94, the part II of “Chinese information processing glossary”: Chinese and Chinese characters. *Zhongguo Biaozhun publishing company*.
- Mingyang Hu, 2002, On foreign letters and foreign abbreviations, *Applied Linguistics*, no. 2. pp. 98-101.
- Yongquan Liu, 1994, Discussing on lettered-words, *Yuwen Jianshe*, no. 10. pp. 21-24.
- Yongquan Liu, 2002, On lettered-words, *Applied Linguistics*, no. 2. pp. 85-90.
- Yunfang Wu, 2003, Studies on modern Chinese parallel structures for Chinese information processing, *Peking University PhD dissertation*.
- ZeZhi Zheng, Pu Zhang, 2005, The Research on Lettered-word Extraction in Chinese Texts, *Journal of Chinese information*, vol. 19, no. 2, pp. 78-85.
- Dictionary compilation room, the language institute, academy of social sciences, 1985, The modern Chinese dictionary, Beijing: *Commerce publishing company*.