Zhang Zhoucai
Research Center of Computer &
Microelectronics Industry Development
China

Vertical Unification of CJK Ideographs

1. Introduction and Definitions

1.1 Horizontal Unification of CJK Ideographs

Horizontal unification of CJK (Chinese-Japanese-Korean) ideographs refers to the task that has been done by the CJK-JRG according to its Unification rules, the output of which is the CJK Unified Ideographs arranged in the Basic Multilingual Plane of ISO-IEC 10646.

The Chinese Hanzi/Japanese Kanji/Korean Hanja are listed in four columns in the code table. If unified, then they have the same ISO code and appear in the same horizontal line.

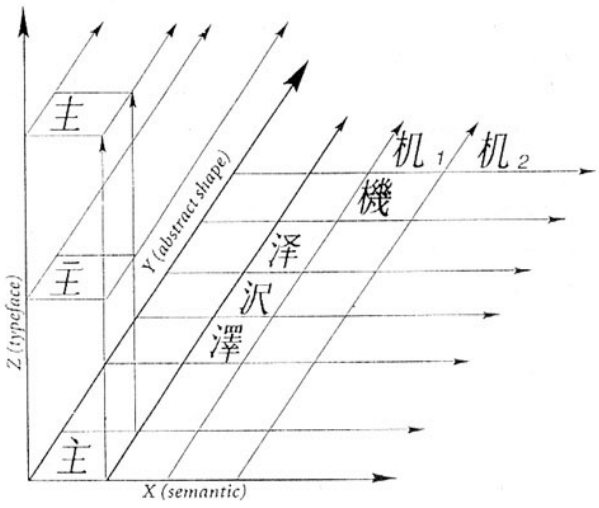
In other words, the above mentioned unification is based on the so-called X/Y/Z model, where *meanings* (the Signified) are arranged along the X axis, *glyphs* (the Signifier: character shapes) along the Y axis, and variants along the Z axis. The unification achieved in the ISO code was performed along the Y axis. This type of unification is referred to as Horizontal Unification (HU).

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

ISO HexCode	C G-Hanzi-T	J Kanji	K Hanja	
4E00	一	一	一	
4E01	丁	丁	丁	
:				
4E0E	与	与	与	←Horizontal Unification
:				
4E30	丰	丰	丰	丰 ←
:				
58F9	壹	壹	壹	壹
:				
8207	與	與	與	與
:				
8C50	豐	豐	豐	
:				
91D8	釘	釘	釘	釘
:				
9488	钉			
:				

As shown in the codetable, HU is done mainly but not absolutely by shape. On the one hand, even some ideographs which have the same meaning(s) and common shapes, showing only minor difference in the direction of individual strokes (such as 丢 and 丟) are *not* unified due to the principle of Source Separation in the CJK Unification Rules; on the other hand, non-cognate ideographs are not unified either, although they may have quite similar shapes (such as 士 and 土). Apparently the meanings of the ideographs were taken into consideration while HU was being done.

In fact, although HU stresses distinctions between glyphs, the horizontally unified CJK ideographs predominately signify the same or related meaning(s), with only a very few exeptions.

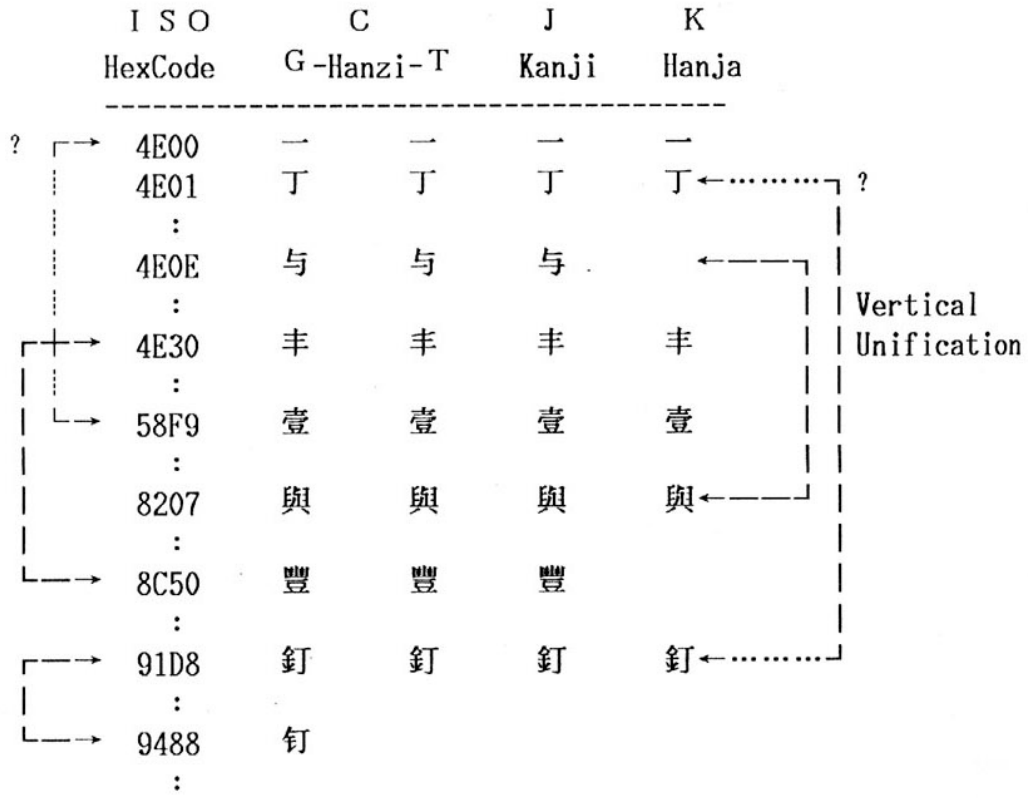


Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

1.2 Vertical Unification of CJK ideographs.

The vertical unification (VU) of CJK ideographs is defined as the one which is based on identity in meaning among ISO-IEC 10646 ideographs, in contrast to HU, which is primarily based on shape.

VU may also be considered as an extension and complement to HU. On the code table, VU would be rendered as upward/downward pointers linking different codes to each other, as shown below:



VU should be performed so as to link characters of identical meaning, but this must be kept within limits. To unify 一 and 壱, which both indicate the number "one", would be a kind of overkill. While the unification of 丁 and 釘 might be linguistically reasonable for some languages, it would be technically unnecessary.

One thing which must be emphasized is that VU does not involve any re-unification or re-encoding of CJK ideographs. All further unification will be based on the output of HU as reflected in the ISO standard, with no further code change.

2. Scope and Limitations of VU

As discussed in this paper, VU might be applicable for ISO-IEC 10646-1 CJK ideographs in applications for glyph normalization and/or the convergence of character

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

repertoires within particular language contexts, or conversion/translation among different language contexts.

To facilitate the discussion and to focus on the most useful CJK ideographs in information technology, this paper takes a subset of CJK ideographs that covers the ideographic characters in the following source sets:

(1) G0	GB 2312-80	used in the PRC (mainland China)
(2) T1	TCA-CNS11643/1st Plane	used in Taiwan
(3) T2	TCA-CNS11643/2nd Plane	used in Taiwan
(4) J0	JIS x 0208-90	used in Japan
(5) K0	KS C 5601-87	used in Korea

As a result of subsetting, the matrix to be dealt with will become more sparse, as shown below:

ISO HexCode	C G-Hanzi-T		J Kanji	K Hanja
4E00	一	一	一	一
4E01	丁	丁	丁	丁
:				
4E0E	与	与	与	
:				
4E30	丰	丰	□	
:				
58F9	壹	壹	壹	壹
:				
8207	□	與	與	與
:				
8C50	□	豐	豐	
:				
91D8	□	釘	釘	釘
:				
9488	钉			
:				

Note: The □ means that the ideograph is out of the subset.

Moreover, in many cases, the characters in K0 are the same as the ones in T1 or T2, so the K column is omitted in some diagrams for simplicity. G, T, J, K are used as abbreviations for these character sets.

3. Classification of Vertical Unification

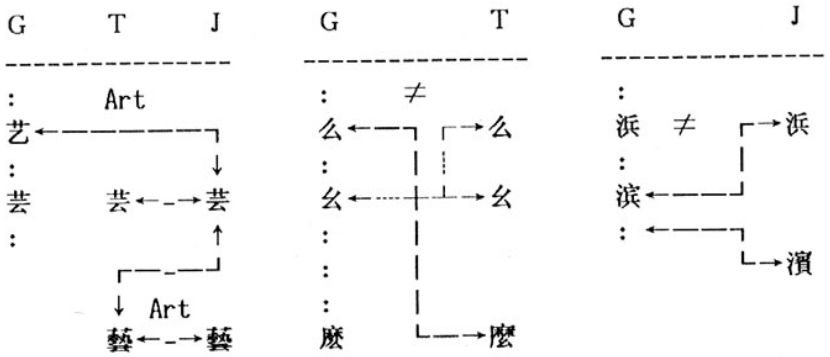
3.1 Homonyms

A few horizontally unified CJK ideographs need to be separated vertically by their meanings. In general, VU is intended to link ideographs or different shapes according to

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

their meaning. A few exceptions, however, result in splitting instead of unification of CJK ideographs. Below are some typical examples. In the first example the simplified character used in Japan with the meaning of "art" happens to be the simplified form of a very different character as used in mainland China.

It should be emphasized again that such cases constitute only an extremely small percentage of the characters in the CJK repertoire; they are however among the most frequently used characters in the modern CJK languages, and must therefore be accounted for.



3.2 Z-Variants

The term Z-Variants refers to those ideographs that have identical meanings and the same generic shape, but that are encoded in different cells in ISO 10646 because they had existed in a source character set as separate code points. The principle of Source Separation meant that this separation had to be reflected in ISO 10646 as well. We may conceive of the Z-Variants as those characters which would have been unified in the process of HU if generic shape had been the only consideration.

3.2.1 Pure Z-Variants

Pure Z-Variants are those variants which have the same number of strokes. Some examples are given below, but this is not an exhaustive list. Z-Variants of this kind are the main ones to cause confusion in the use of software applications, and may cause failure in text searching and retrieval if care is not taken.

兑 丢 尔 内 凛 刊 别 呐 囱 匀 吞 册 画
兌 丢 尔 内 凛 刋 别 呐 囪 勻 吞 冊 画

娱 户 彦 悦 术 查 毁 秃 税 脱 蜕 禀 吴
娛 戶 彥 悅 朮 查 毀 禿 稅 脫 蛻 稟 吳

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

3.2.2 Quasi Z-Variants

Quasi Z-Variants are those variants which have the same generic shape, but differ in number of strokes.

并	冲	强	凉	吕	换	刹	厦	奥	粤	抛	厢	刊	对	压	隆	厅	兹
並	沖	強	涼	呂	換	剎	廈	奧	粵	拋	廂	刋	対	圧	隆	庁	茲

3.3 Orthographic and Variant Ideographs

Depending on the language context, an ideograph viewed in one country or region as an orthographic ideograph (standard character) might be viewed as merely a variant ideograph in another country or region. In a multilingual and international environment, it is very difficult to determine whether or not an ideograph is an orthographic one or not due to political and/or cultural reasons. The following examples show orthographic ideographs in G (which would be viewed as variants in T and J) with their corresponding orthographic ideographs in T and J (considered to be variants in G). It can be seen that orthographic ideographs and variants, and Z-Variants overlap each other.

G：	并	采	耻	冲	吊	仿	挂	炮	异	游	韵	咏	烟	灾	册
	|	|	|	|	|	|	|	|	|	|	|	|	|	|
JT：	並	採	恥	沖	弔	倣	掛	砲	異	遊	韻	詠	煙	災	冊

3.4 Old and New Glyphs

The distinctions between old and new glyphs is similar to that between simplified and unsimplified glyphs discussed in 3.5 below, in that "old" and "new" are relative concepts influenced by political and cultural factors, and are very difficult to deal with. Note that the following examples are seen as distinctions between old and new forms, not as distinctions between simplified and unsimplified forms (seen from the point of view of a Chinese):

吕	侣	唤	换	吴	宫
|	|	|	|	|	|
呂	侶	喚	換	吳	宮

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

3.5 Simplified and Unsimplified Ideographs

3.5.1 One-to-One Relations

In the unified CJK ideographs, over 2,000 CJK ideographs have such relations between each other, which may be denoted as 1:1. Almost all simplified ideographs have unsimplified counterparts. The reverse is not true. That is, many unsimplified ideographs do not have corresponding simplified forms in the repertoire.

Simplified Hanzi/Kanji versus unsimplified ones:

宝	蚕	痴	点	独	断	国	会	号	旧	礼	猫	声	区	双
寶	蠶	癡	點	獨	斷	國	會	號	舊	禮	貓	聲	區	雙

Hanzi and Kanji simplified in different ways:

G	J	G	J	G	J	G	J
对	對	对	厅	廳	厅	单	單	单	画	畫	画
压	壓	压	辩	辯	弁	证	證	証	团	團	団
边	邊	辺	带	帶	帯	举	舉	挙	弹	彈	弾
齿	齒	歯	处	處	処	恶	惡	悪	发	發	発
丰	豐	豊	广	廣	広	继	繼	継	两	兩	両
觉	覺	覚	济	濟	済	齐	齊	斉	荣	榮	栄
气	氣	気	卖	賣	売	脑	腦	脳	实	實	実
图	圖	図	围	圍	囲	释	釋	釈	泽	澤	沢
亚	亞	亜	应	應	応	营	營	営	县	縣	県
传	傳	伝	价	價	価	从	從	従	听	聽	聴

Simplified Hanzi only, no simplified Kanji exist:

说	纺	饭	锻	车	轴	鸟	鹏	风	麦	麸	鱼	爱	币	飞	汤
說	紡	飯	鍛	車	軸	鳥	鵬	風	麥	麩	魚	愛	幣	飛	湯

Simplified Kanji only, no simplified Hanzi exist:

州	知	予	炎	希	黒	毒	徳	喝	突	仏	抜	壱	仮	隆	増
洲	智	豫	焔	稀	黑	毒	德	喝	突	佛	拔	壹	假	隆	增

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

3.5.2 One-to-Many Relations

There are two kinds of one-to-many relations. In one, an ideograph in a certain context corresponds to more than one in another context, (denoted as 1:M). In the examples below, the simplified forms correspond to distinct unsimplified forms:

开 頭
发：發 － 髮

钟：鐘 － 鍾
錶 情

In the other kind of one-to-many relation, an ideograph in a certain context corresponds to more than one ideographs, including itself, in another context (denoted as 1:M+1). In the examples below, a pre-existing ideograph is used in its original use, and as a simplified version of one or more other characters.

前 皇
后：後 － 后

注：注 － 註
意 解

樹 若
干：榦 － 乾 － 幹 － 干
淨 部

里：里 － 裡 － 裏
程 面 面

制：制 － 製
度 造

3.6 Proposed VU Classification

Among the approaches noted above for classifying ideographs which may be unified vertically, we may propose overlap and dependency. In order to eliminate or reduce the dependency of VU classification, from the viewpoint of the X/Y/Z model the VU objects should be organized in the proposed frame. From the viewpoint of a software developer, the relational attributes 1:1, 1:M and 1:M+1 make more sense than the other attributes for defining the data structure of mapping tables.

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

		1:1	1:M 1:M+1
Z	Pure Z_Variants	丢／丟 兑／兌	
Y	①Quasi Z_Variants	吕／呂 冲／沖	注／注註
	②Y_Variants	奥／奧 厢／廂	
	③Simplified Unsimplfd.	韵／韻 烟／煙	
		断／斷 独／獨	发／發髮
		贝／貝 鱼／魚	后／後后
		旧／舊 宝／寶	钟／鐘鍾
X	X_Variants = Homonyms	芸／芸 浜／浜	

4. Characteristics of VU

4.1 Language Context Dependency

The results of VU depend on which language contexts are related. the examples below show how different it is when VU is done between GT (mainland China and Taiwan) and between GJ (mainland China and Japan). For example, the character 連 in modern Japanese is customarily used as an equivalent for 聯, so both of these must be related to the mainland Chinese character which corresponds only to 聯 in Taiwan.

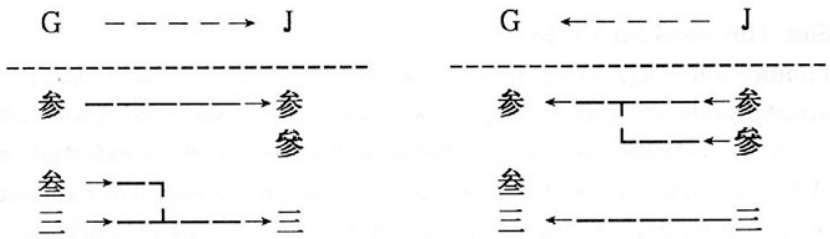
G T	G J
联 — 聯	联 — 聯
	＼
连 — 連	连 — 連
讽 — 諷	讽 — 諷
	＼
风 — 風	风 — 風
诀 — 訣	诀 — 訣
	＼
决 — 決	决 — 決

4.2 Direction Dependency

The results of VU also depend on the direction along which the vertically unifiable ideograph points to its counterpart from one context to another. The examples below show how different it is when VU is done from G to J, versus from J to G. In this case links are

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

drawn so as not to point to characters which are not in standard use, but which exist as mere variants.



4.3 Coupling Intensity

The tie between ideographs linked by VU has an "intensity" which reflects the necessity, possibility and frequency of a link to one ideograph as compared to possible links to others. The coupling intensity also reflects contexts and direction. For example, the coupling intensities between 一 and 壹 may be set to zero due to unnecessity; from G to J, the link from 叁 to 三 may be set to greater than zero, while the corresponding link from 三 to 叁 may be set to zero, since the latter character is not in standard use. Further, for the links from 几 in G to 幾 and 几 in J, the first one will have a greater intensity than the second, which may be set according to statistics from a corpus.

5. Case study: an Implementation of VU

One of the applications for VU is to convert text files between (Chinese) simplified context to unsimplified context. A conversion software has been developed by Eten Information System Co. It is called BMT for "The Bridge between Mainland and Taiwan". BMT enables the coexistence of simplified/unsimplified/variant Hanzi in a single computer system, the bi-directional conversion between GBcode and Big5, as well as the (1:M)/(1:M+1) handling.

The implementation of the BMT indicates that the classification of VU is preferable to be done in terms of (1:1)/(1:M) or (1:M+1). When orthographic characters and variants; new and old glyphs; and simplified and unsimplified characters are mixed together, more non-one-to-one relations are derived, which made the process more complicated. A special routine with a sophisticated data structure handles each mapping relation. For non-one-to-one relations, the unified candidates are properly ordered, the one with the higher frequency chosen first, then checked and corrected by users. The (1:M)/(1:M+1) relations not only exist from G to T, but also appear from T to G. For example, 著 in T has two counterparts, 着 and 著 in G which appear very often in text converting. Some unsimplified Hanzi (such as (麽 and 後) in G makes confusion likely. Thus it would be wiser to ignore them and adopt the simplified ones (么 and 后). In general, G contexts use

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

traditional and simplified Hanzi only, which differes from J contexts in which mixed simplified and unsimplified Kanji are allowed.

6. Conclusion: The Need for VU Rules

Vertical unification of CJK ideographs is a significant and complicated task. It calls for the close collaboration of Chinese, Japanese and Korean scholars. Like horizontal unification, vertical unification needs rules before it may be formally conducted. Some of the factors discussed in this paper must be taken into account to establish reasonable and workable rules. In particular, the classification and quantification of VU attributes are the first things that must be determined. Accordingly, the following six pairs of tables must be developed:

G → T / T → G
G → J / J → G
G → K / K → G
T → J / J → T
T → K / K → T
J → K / K → J

As a result, a new database is expected to be organized and shared by users of CJK ideographs, to promote the multilingual exchange of culture, science and technology.

ＣＪＫ（中日韓）統合漢字の垂直統合

（要約）

1．概論と定義

1.1 ＣＪＫ漢字の水平統合

ＩＳＯや中国，日本，韓国の漢字統合は，意味より字形からとらえている。Ｘ軸に意味，Ｙ軸に字形，Ｚ軸に変種というＸＹＺ座標系に基づき，同じ漢字はＹ軸を中心に水平統合されている。しかし，中国，台湾，日本や韓国の漢字の互換領域を考えるとき，字形重視の水平統合は，言語的文脈や簡化による差異，類義，多義性などの問題に対処できない。

1.2 ＣＪＫ漢字の垂直統合

そこで，ＣＪＫ漢字を，字形重視のＩＳＯ，水平統合における意味を柱に，1.2 の表のような垂直統合により定義つけてみる。

2．垂直統合の適用範囲と限界

ＩＳＯ，ＣＪＫ漢字表に，字形の標準化，異言語間での変換，言語的文脈による漢字のレパートリの適用による垂直統合を行ってみる。漢字を中国の GB 2312-80，台湾の TCA-CNS11643，日本の

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

JIS x 0208-90, 韓国の KS C 5601-87 という各国の漢字表をソース・セットとしてＣＪＫ漢字表のサブ・セットを考えると，ＩＳＯとの間で取り扱われている漢字に差（２の表の四角部分）がある。

3. 垂直統合の分類

3.1 同形異義字

水平統合によるＣＪＫ漢字統合は，意味による垂直統合では表のような漢字間の分離が生じる。それだけに意味による関係づけが必要と言える。

3.2 Ｚ変体

ソース・セパレーションによりＩＳＯではコード化が異なるが，同じ字形で意味も同じもの（例：冊－册）。

3.2.1 純Ｚ変体

同じ画数のもの（例：稟－禀）。

3.2.2 準Ｚ変体

画数に一つか二つの違いがあるもの（例：冲－沖）。

3.3 正字と異体字

国や地域，政策や文化的側面により正字が異なる。例字は中国と台湾における正体字である。正字／異体字とＺ変体とに重なりが見受けられる。（例：咏－詠）

3.4 新字形と旧

3.5 の簡化に似てはいるが，新旧字形は政策，文化などと関連するものであり，漢字の簡化や中国における簡略化とは異なるものである。（例：呂－呂）

3.5 簡体字と繁化字

3.5.1 １対１の関係

ＣＪＫ漢字統合において，約２千のＣＪＫ漢字が１対１の関係にある。ほとんどの簡化文字は繁化文字と対応するが，その逆は言えない。（例：希－稀）

3.5.2 １対複数の関係

これには２種類ある。(1) ある文脈で使われる漢字が，ほかの文脈で使われる複数の漢字に対応する。「１対Ｍ」（例：发－發－髮）

(2) ある文脈で使われる漢字が，ほかの文脈で使われる同じ字や他の字に対応する。「１対Ｍ＋１」（例：里－裡－裏）

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

3.6 垂直統合分類案

垂直統合による分類を具体的に行うことによって漢字の重なり具合や依存度が分かる。そこで，垂直統合から漢字の依存度を下げたり取り除くために，１対１，１対M，１対M＋１の属性を加えたＸＹＺモデルを用いて，垂直統合のモデル図を作成した。

4. 垂直統合の漢字

4.1 言語的文脈依存

垂直統合は，台湾と中国，中国と日本の例に示されるように，言語的文脈に依存する。

4.2 統合方向依存

垂直統合は，中国から日本と日本から中国の例に示されるように，ある言語的文脈からある文脈へ統合する方向に依存する。

4.3 漢字間の結合は，漢字間における結合の必要性，可能性，頻度の度合いによる。また，言語的文脈や統合方向にも関係する。

5. ケーススタディー垂直統合の実行

垂直統合の対象の一つは，中国語と台湾語間におけるテキスト・ファイル変換である。一つのコンピュータ・システムで簡体字・繁字・異字を取り扱え，１対M／M＋１処理，ＧＢコードとBig 5間の双方向コンバートする変換ソフトＢＭＴ（Bridge between Mainland and Taiwan）を使い，１対１と１対M，あるいはM＋１に関して垂直統合の分類を実行した。正字・異体字，新字形・旧字形と簡体字・繁字が混在すると，無対応が起こり，複数なデータ構造のルーティンにより関係づけが処理された。

6. 結論：垂直統合規格の必要性

漢字統合のためには垂直統合は重要かつ複雑な作業であり，垂直統合の分類やその属性などを検討し，漢字統合の規格を作り，中国，日本，韓国間の漢字コードを開発する。そして，ユーザーにより新しいデータベースが構築され，使用されていくことが望まれる。

コメンテータ：松岡栄志（東京教育大学）

最初に，ISO/IEC 10646 を巡って最近の動向を説明したい。ISO/IEC 10646 は，世界中の全ての漢字と記号を一つのコードに収める目的で作られた。昨年５月に ISO(国際標準化機構) で承認された。ISO/IEC 10646 は，中国，日本，韓国をはじめ漢字を使っている国々が話し合い「ユニフケーション」することになった。実際の作業の中心的な働きをしたのが張先生です。日本側は，私が委員として出ました。ユニフケーションは，例えば「言」という字は，中国と日本では鍋・蓋の点の方向が違う。これをユニファイしてどうかと言うことが問題になる。日本は，最初ユニフケーションに反対であったが，最終的には

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

CJK-JRK ジョイント・リサーチ・グループに参加し検討に加わった。先程の「戸」は分け，「言」はユニファイした。それが微小な差だと言うことだが，本当にそうか問題が残る。張さんには，どう捉えるのか質問したい。

最初の検討は，全体の規則を検討したがソースセパレーションと XYZ モデルをやった。各国で分けているものは分け，分けないものは統合するということで，中国・台湾・日本・韓国という順序で横一列に配列した。その結果，日本の JIS X0208 や JIS 0212，中国の簡体字，繁体字(旧字) 韓国の漢字をあわせて 20,902 種をコード化した。実際の運用については，幾つかの問題があると思う。その一つは，検索の問題である。コード配列は，康熙辞典の部首を使った。部首は，総数 1 万を越えると検索が大変になる。そのため，漢字属性辞典が必要になる。特に，国際版では早急に必要だ。中国では，すでに出ているが，日本の音・訓情報が入っていない。

次の質問は，中国語のデータベース化について 20,902 字では足りない問題がある。私達の CJK-JRG グループを新しく IRG というイディオグラフィック・ラボタ・グループと改称し，拡張について議論を進めている。この企画は，まもなく JIS において国際企画として認められる方向にある。その場合，注意が必要なのは，現在日本で使われている JIS コードと違うコードが与えられている。次に，字形の問題がある。これは，意味の分類を含めて共同の研究が必要である。これまでは，ユニフケーションが意味を除外していたが，意味を入れることは文字一字一字について研究が必要だ。もう一つは，自分の国以外の情報処理の状態や文字の研究が知られていない。やはり，交流とか情報の受け渡しをやるべきだ。最後に，ISO/IEC 10646 の実用化の時期について張さんはどう思うか伺いたい。

Zhang: Regarding vertical unification, following the suggestion of an American linguist, I now believe that identification rather than unification is a more workable concept.

One thing that must be understood is that vertical unification does not involve any change in the actual codes.

Another point is that vertical unification is extremely complicated. It is important not to have too much, or too little unification.

The job of vertical unification should be performed by academic people rather than by politicians. The results of vertical unification are to provide a reference, not to implement a standard.

It is difficult to say when ISO 10646 will be implemented. In China and Japan there is no possibility of developing our own operating systems. We have been working, however, with people from Apple, Windows, Taligent and Sun Systems on the implementation.

There will still be work for us to do in Asia, especially on the vertical identification of ideographs, which is urgently needed to help the exchange of information in multilingual environments such as GATT, WTO and APEC.

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

I cannot say when implementation will come, but I feel that it will be soon, say in a couple of years.

宮沢彰（学術情報センター）：学術情報センターの宮沢と申します。バーティカル・ユニフケーションについてコメントしたい。これまでのホリゾンタル・ユニフケーションは，スクリプトのレベルで処理していた。バーティカル・ユニフケーションは，明らかに意味に係わる部分を含んでいる。意味は，文字だけではなく言語のテキストをもってきてはじめて確立できる。本来，日本語の中で使われているものと，中国語の中で使われているものと，どの様に対応づけられるのか。一方，漢字は，イディオグラフィックであると言うことで文字レベルで同じだ。3ページの例で，4E0E と 8207 では，私しは同じと判断するが，人によって違うこともある。これは，色々のレベルが混在しているために，この作業を進めると大変になる。これを避けるためには，目的をはっきりさせ，例えば日本語と韓国語の翻訳に必要な対応表，目録の情報交換に字体変換なしにもとの形で交換するために目的をはっきりさせて作業を進めることが必要だ。例えば，我々だと，J から J へのバーティカル・ユニフケーションも必要だ。「さわ」が旧字であっても，新字であっても「みやざわ」を検索するさい一緒に出て欲しい。これをはっきりさせれば充分意味があると思う。

Zhang: I agree that the identification is needed within a one-language environment, for normalization purposes. There is no copyright issue involved. The output of the identification process will be a reference data table, not a standard.

The vertical unification will only define the possibility of relationships, leaving the right to choose up to the users. The actual mapping that is used will depend on the environment or context.

米田：先程，松岡先生から検索についてのご質問がありましたが，ご回答はいかがでしょうか。

Zhang: Vertical identification will make text retrieval easier, improving the hit rate—once it is implemented, not in existing systems.

Sharpe: Is there more similarity between Chinese and Japanese or Korean characters in terminology from specialized fields, such as law or medicine, than in basic terms, that is the core vocabulary? Would it be better to first apply the system to fields in which there is more similarity, and then extend it to other fields later?

宮沢彰：そのような用語になると一般的に二字以上の熟語が多くなると思う。このレベルで言っていいるのは，一字一字のレベルを想定していると思う。そういうターミノロジーに対応表は，レベルが言語に近い，上のものになると思う。

松岡：ホリゾンタル・ユニフケーションは，「形」だけでいけるという話だが，さきほどの「言」「言葉」「戸」という字が，「言葉」と「戸」は一緒だけど「戸」が分けてあるという問題がある。日本人がこの表を見ると矛盾がある。矛盾は，意味を排除して字の

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层

事情を排除して「形」だけでやった結果だと思う。張さんの垂直統合は，この補充という形が出てきたと思う。ですから，形だけでいけるかどうかが一番大きな問題だ。

ただ，今お話になった，その例について言えば。それは，ソース・コード・セパレイションの原則ですね。

松岡：ええそうです。

宮沢：それは，ご指摘の理由ではないと思う。一般的に松岡先生の疑問については，私もそうだと思う。ただ，具体例はちょっと違うと思う。

當山日出夫（花園大学）：垂直統合とおっしゃる内容は，理解できる。それを，日本で人文科学，中国文学とか日本文学，日本史をやっている人間が，普通使う用語に直せば，漢字シソーラスという用語になる。現在の JIS コードのなかで新字体と旧字体を統合し変換したい。あるいは，データベースを作るときにどちらに統一するかと言う議論はある。ISO/IEC 10646 が実現した場合にそれに対してどう対応するかと言う現実的な問題だと思う。実際の日本語の研究者に，どう提案するかとなれば，漢字シソーラスという用語がなじみやすい。

それから，昨日のシンポジウムで話しのあった，文字の問題は，純粋に学術的に考える部分と，文字・言語政策という部分がある。この場合，必要なのは，国立国語研究所などの調査で，漢字の字体を我々と外国人の意識は，どうかと言う基礎的な調査も必要だ。調査なしで，お互いの感想だけで言っていいると水掛け論になる。

林 大（評議員）：ISO コードは，日本と中国と韓国の間で情報交換を円滑にするために行うために必要だと思う。それは，印刷文献の問題で，日本人がこれを聞くと，手書きの問題と混同する。仕事は，印刷物が普通だ。それぞれの国で印刷の形についての問題であると言うことをはっきりさせるべきだ。すぐにこれを批判すると手書きの問題に入る。昔は，「松」は「木」の下に「公」とかいたり，「公」を「船」の形に「ハ，クチ」と書いたりする，などと言う。ここでの問題は，今は，活字にあるものについの問題だと言うことが必要だ。後は，今おっしゃった漢字シソーラスと言うことになったら，日本での手書きの昔からの形も，どの様に記述するかと言う問題も含めて考える必要がある。

Zhang: My understanding is that the main purpose of the proposals under consideration are concerned with information exchange, such as over Internet, not with the printing of texts.

Contact: Zhang Zhoucai, email : joe.zhang@unihan.com.cn TEL: 86-10-6234 5536 FAX: 86-10-6231 2924
Address: No.14, 3/F, ZhiXinCun XiaoQu, Haidian District, Beijing 100083, China 北京志新村小区 14 号 3 层