

PINYIN CONVERSION SPECIFICATIONS FOR NAME AUTHORITY RECORDS

CONVERSION OF MIXED TEXT (ISWG)

OCTOBER 1, 2000

1 Instructions

- 1.1 The programs that apply the Mixed Text conversion sequences use ApplyIWG to count each type of syllable and evaluate the counts to determine what value to set for the subfield for IsWadeGiles' evaluation in the Subfield→Status array.

NOTE: Each subfield has an array. This array contains a status value for its evaluation by each conversion sequence that has evaluated it. So each subfield has a status, such as DoNotConvert or NoChanges, for each conversion sequence.

- 1.2 Before evaluating the counts, ApplyIWG's preliminary processing:

- ◆ identifies subfields containing “Taiwan))” or “.,Taiwan)” or English words such as “College”, “University”, “Society”, etc.
- ◆ identifies matches to the Archaic list and blocks matching text from being converted using a blocking array. The blocked array has the same number of characters as a subfield, and indicates those characters that are to be ignored during conversion.
- ◆ checks the Stoplist for a match. If it finds a match to both before and after, it sets (IWG) status (in status array) to DoNotConvert. If it finds a match to the before version only, then it sets the status to NoChanges.
- ◆ checks C1 and then C2 to see if the subfield was already converted. If so, it sets the status to DoNotConvert.
- ◆ checks tokens for abbreviations and non-text characters. If an abbreviation is found the subfield is marked DoNotConvert.
- ◆ checks for “too common” syllables, i.e., those that are found in other languages as well as the Wade-Giles format of Chinese. (This list refines the rules in the table.)
- ◆ checks for suffixed syllables (xiang, xian, zhou, qu and qi) to try to allow for already converted text that would not preclude conversion.

- 1.3 The structured English that follows describes the rules that ApplyIWG uses to evaluate subfields (or subsections) and the status that is set if a subfield meets the listed conditions.

- 1.4 The following abbreviations are used in the structured English. **The categories of syllables are laid out in the Data Dictionary SYLLABLE: five categories of syllables.**

The category designated...	contains syllables that are...
WG	unique to the Wade-Giles format.
Pinyin	unique to the Pinyin format.
Same	not changed when translated from Wade-Giles to Pinyin; the same in both formats.
Common	found in both formats, but not as a direct translation.
Other	on the Stopword list, a non-text string such as “1949-“, or other unaccounted-for non-Chinese text.

- 1.5 For all tables in the structured English, rows are joined by a logical “or”; if a subfield meets any row’s conditions, then a decision is made and the evaluation goes no further. Order of evaluation is therefore significant.

2 Conversion action

- 2.1 IF subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
only Same	-
only Same and Other	-
Only Same and non-text	-
Same + words >= cnttok	-

THEN there are **no Chinese syllables, or syllables will convert to be the same**. The status for IsWadeGiles in the Subfield → Status array for this subfield is set to **DoNotConvert**.

DONE.

OTHERWISE

- 2.2 IF subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
there is only one syllable in the wg, same, and common	-

categories combined and it is Too Common	
------------------------------------------	--

THEN the only convertible syllable is Too Common. Set status to DoNotConvert.

DONE.

OTHERWISE

2.3 If subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
same and common	wg and pinyin
common	wg and same and pinyin and other
same and common and other	wg and pinyin

THEN the text contains common or common and same syllables and must be reviewed. The status is set to ReviewRequired.

DONE.

OTHERWISE

2.4 If subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
same and common and other and one wg	pinyin

THEN the mixed text must be reviewed. Status is set to ReviewRequired.

DONE.

OTHERWISE

2,5 IF subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
only Pinyin	-
only Other	-

stoplist and non-text	wg and pinyin and common and other
wg==0 and same==0 and comm==0	-
wg==1 and same==0 and comm==0 and toocmn	-
Pinyin	same and common and wg and other
Pinyin or same and college	wg
same and pinyin	wg and common and other
same and pinyin and other	wg and common
same and other	wg and common and pinyin
pinyin and other	same and wg and common
other	same and wg and common and pinyin
same and common and pinyin	wg and other
common and pinyin	same and wg and other
common and pinyin and other	same and wg

THEN **no Wade-Giles was found to convert.** The status is set to DoNotConvert.

DONE.

OTHERWISE

2.6 IF subfield

CONTAINS	AND DOES NOT CONTAIN
wg syllable(s) all of 1 character length	pinyin and same and common
pinyin syllable(s) all of 1 character length	wg and same and common
wg and pinyin syllables all of 1 character length and other	same and common

THEN the only matches are to **one-character syllables**, which are too generic to be considered valid matches. SingleWadeGiles.

DONE.

OTHERWISE

2,7 IF

CONTAINS	AND DOES NOT CONTAIN
wg and the word “College” or “University” or similar words such as “Society”, etc.	-

THEN the string contains both WG and the word “**College**” or “**University**” or “University” or similar words such as “Society” from a list provided by LC.and requires manual review. ReviewRequired.

DONE.

OTHERWISE

2.8 IF

CONTAINS	<u>AND DOES NOT</u> CONTAIN
one syllable total in the wg, same and common categories combined and less than three other syllables	too common and pinyin

THEN there are a few **Other** syllables that should be manually reviewed. ReviewRequired.

DONE.

OTHERWISE

2.9 IF

CONTAINS	<u>AND DOES NOT</u> CONTAIN
one syllable total in the wg, same and common categories combined and three or more other syllables	pinyin
two syllables total in the wg, same and common categories combined	-

and too common and more than three other syllables	
----------------------------------------------------	--

THEN there is **one WG with multiple Other** syllables. These subfields should be **skipped** DoNotConvert.

DONE.

OTHERWISE

2.10 IF subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
only one syllable and it is wg	too common (i.e., the one syllable is not on the too common list)
one wg syllable and non-text characters	anything else
one wg	college and pinyin and common and same and other
only one syllable and it is common	-

THEN a **single Wade-Giles** syllable was found. It is **not converted**. Subfield→Status is set to SingleWG.

DONE.

OTHERWISE

2.11 IF subfield

CONTAINS	<u>AND DOES NOT</u> CONTAIN
Common and more than one wg and more than one pinyin	-
Common and wg and pinyin	Too common

THEN the **mixed text must be reviewed**. Subfield→Status is set to ReviewRequired. ApplyIWG checks for special Taiwan status also.

DONE.

OTHERWISE

2.12 IF subfield

CONTAINS	<u>AND DOES NOT CONTAIN</u>
pinyin and more than two wg syllables	common and other

THEN the subfield contains **mixed** unique WG and Pinyin syllables. These subfields are **converted**. Subfield→Status is set to NoChanges.

DONE.

Evaluate conditions that would allow conversion.

OTHERWISE

2.13 IF subfield

CONTAINS	<u>AND DOES NOT CONTAIN</u>
wg and common and same	pinyin and other
same and wg	common and pinyin and other
wg	same and common and pinyin and other
wg and common	same & pinyin and other

THEN the subfield should be **converted**. Subfield→Status is set to NoChanges.

DONE.

OTHERWISE

2.14 IF subfield

CONTAINS	<u>AND DOES NOT CONTAIN</u>
Same and pinyin and other and only one wg syllable	-

THEN the subfield should **not be converted because it has only one WG syllable**. Subfield→Status is set to DoNotConvert.

DONE.

OTHERWISE

2.15 IF subfield

CONTAINS	<u>AND DOES NOT CONTAIN</u>
common	same and wg and pinyin and other
common and other	same and wg and pinyin
same and wg and common and other	pinyin
same and wg and pinyin	common and other
same and common and other	pinyin and wg
same and pinyin and other and more than one wg	-

THEN the subfield contains **very mixed text** and a **manual review** is required.
Check for Taiwan special cases. ReviewRequired.

DONE.