

# Mapping characters for PRECIS classes

draft-yoneya-precis-mappings-03

Takahiro NEMOTO

[t.nemo10@kmd.keio.ac.jp](mailto:t.nemo10@kmd.keio.ac.jp)

Keio University / JPRS

# Changes from -02 to -03

- -03 defined characters that local case mapping targets.
- -03 requested creating registry of precis local case mapping to IANA and defined a template for registry of precis local case mapping.
- -03 added the initial precis local case mapping registrations.

# Open issues

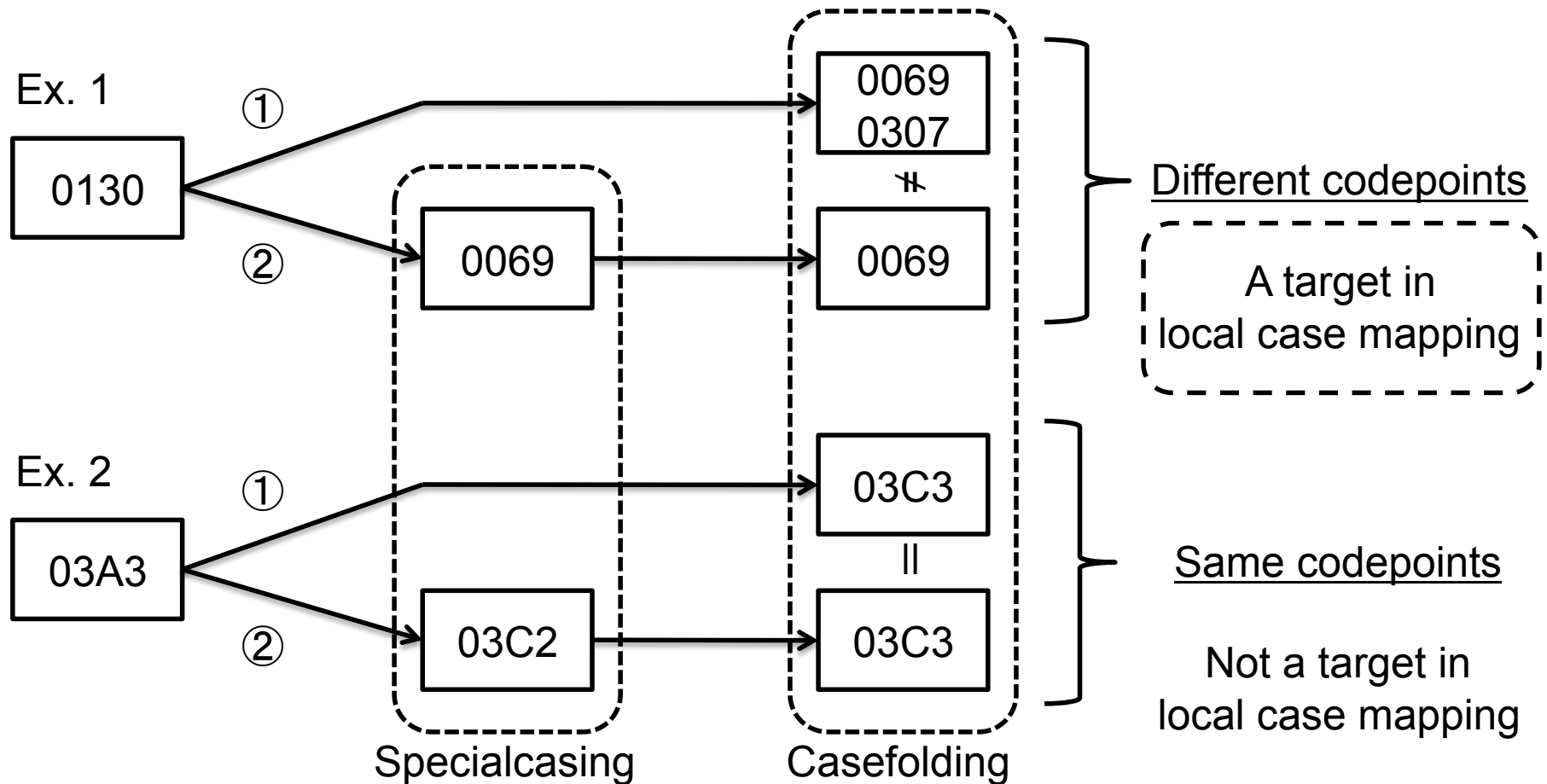
1. Are target characters of local case mapping OK?
2. I'd like to confirm the mappings order.
  - Additional mapping then Case mapping?
  - or
  - Case mapping then Additional mapping?
3. Is an IANA template for registry of precis local case mapping OK?

# **ISSUE 1: TARGET CHARACTERS OF LOCAL CASE MAPPING**

# Local case mapping definition

- Local case mapping is case folding that depend on language context and one of protocol dependent mapping.
- Characters that local case mapping targets are defined based on Specialcasing.txt in section 3.13 of The Unicode Standard.
  - But, our I-D omits characters, that obtain the same codepoint(s) from performing just casefolding and local case mapping then casefolding, from this target.
- we checked all unicode characters that local case mapping targets with specialcasing.txt of unicode standard.

# Example checking characters



Type of processing order: ① Only casefolding ② Local case mapping then casefolding

# A list of characters that local case mapping targets

- **Lithuanian**

**I** = 0049; 0069 0307; 0049; 0049; It More\_Above; # LATIN CAPITAL LETTER I

**J** = 004A; 006A 0307; 004A; 004A; It More\_Above; # LATIN CAPITAL LETTER J

**Į** = 012E; 012F 0307; 012E; 012E; It More\_Above; # LATIN CAPITAL LETTER I WITH OGONEK

**Ì** = 00CC; 0069 0307 0300; 00CC; 00CC; It; # LATIN CAPITAL LETTER I WITH GRAVE

**Í** = 00CD; 0069 0307 0301; 00CD; 00CD; It; # LATIN CAPITAL LETTER I WITH ACUTE

**Ĩ** = 0128; 0069 0307 0303; 0128; 0128; It; # LATIN CAPITAL LETTER I WITH TILDE

- **Turkish, Azeri**

**İ** = 0130; 0069; 0130; 0130; tr; # LATIN CAPITAL LETTER I WITH DOT ABOVE

**Ĭ** = 0130; 0069; 0130; 0130; az; # LATIN CAPITAL LETTER I WITH DOT ABOVE

**I** = 0049; 0131; 0049; 0049; tr Not\_Before\_Dot; # LATIN CAPITAL LETTER I

**I** = 0049; 0131; 0049; 0049; az Not\_Before\_Dot; # LATIN CAPITAL LETTER I

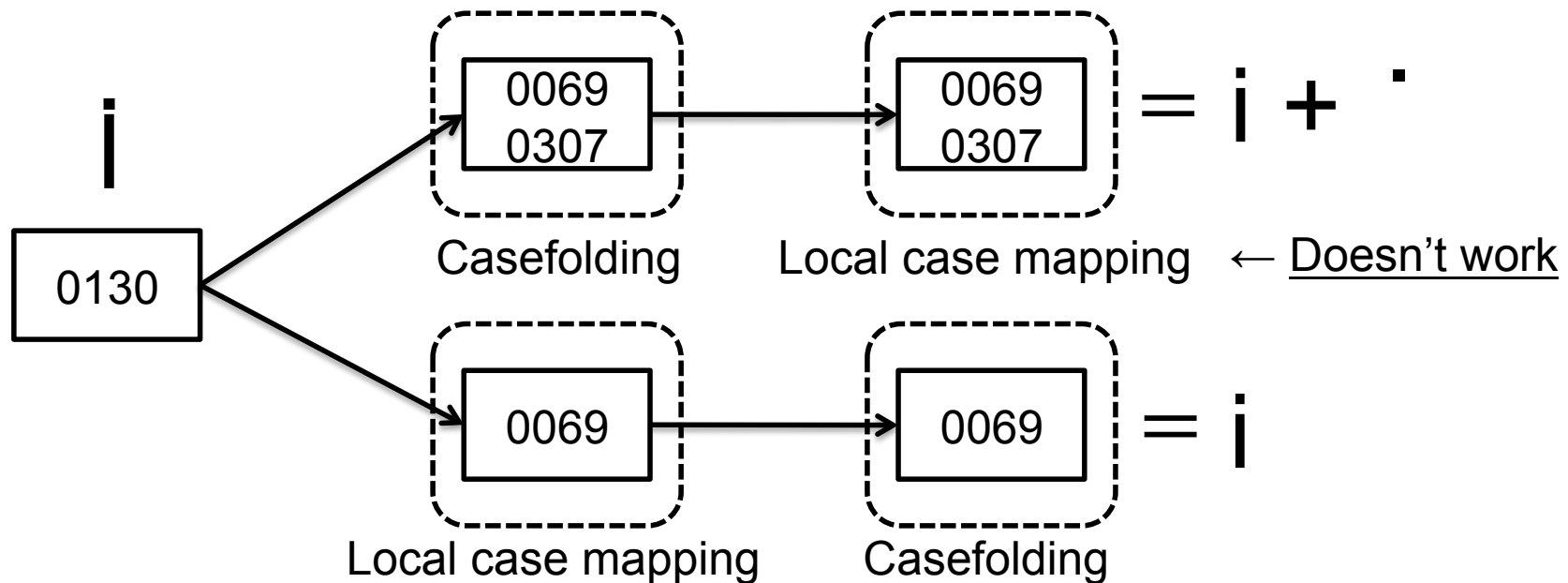
- These mappings are defined in subsection “Language-Sensitive Mappings” in section “Conditional mappings” in Specialcasing.txt

# **ISSUE 2: THE MAPPINGS ORDER**



# Mappings order

- Precis-framework-06 defines process order as casefolding then additional mapping. If additional mapping may include local case mapping, it's preferable to perform additional mappings then Casemapping.



# **ISSUE 3: AN IANA TEMPLATE FOR REGISTRY OF PRECIS LOCAL CASE MAPPING**

# The registration to IANA template

- Followings are the registration to IANA template for precis local case mapping;

Language: Language name

Codepoint: Local case mapping that can be applied  
when this code point exists in the strings

Local lowercase: The lowercase codepoint after  
performing local case mapping

Comment: Character name of the code point

# Example initial registration

language: Turkish

Codepoint: U+0130

Local lowercase: U+0069

Comment: LATIN CAPITAL LETTER I WITH DOT ABOVE

Codepoint: U+0049

Local lowercase: U+0131

Comment: LATIN CAPITAL LETTER I

# NEXT STEP

- Wants to move this version to WG I-D
  - Any comments before adoption?