

2015年2月4日
JGP会合(6)
資料5

日本語LGRの検討課題と 解決方針

日本語LGRで決めるべきこと

- a. 文字範囲(Repertoire)
- b. 異体字(Variants)
- c. 処置(Disposition) (*1)
- d. ラベル評価ルール(Whole Label Evaluation (WLE) Rules) (*2)

(*1) Rootゾーンにおける異体字の取り扱いを決めるもので、個々の文字に対して定義される。値には「登録可能」、「ブロック」などがある。

(*2) 個々の文字が組み合わされた文字列(ラベル)の有効性を評価するルールで、例えば中国語では簡体字と繁体字の混在は登録不可能とするなどが定義される。

a. 文字範囲

- 考えられる選択肢

- ① 常用漢字・人名用漢字の範囲

日本国内の地名や歴史的名称を表現するには不十分である。例えば
奈良県橿原市の「橿」はこの範囲に含まれない(約3000文字)

- ② JIS X 0208:2012の第一水準・第二水準の範囲

常用漢字(JIS X 0213:2012の第三水準に含まれる4文字(叱剝填頬)を除く)・人名用漢字を包含し、かつ日本国内の地名や歴史的名称を表記でき、JPドメイン名でも有用性が確認されている(文字の多寡について苦情や要望がない)(約6350文字)

- ③ JIS X 0213:2012の第三水準・第四水準までの範囲

平均的な日本人が通常使用しておらず、また理解していない(読み書きできない)第三水準・第四水準の文字を入れるのはリスクが大きい(約10050文字)

- ④ IICORE(*)の範囲

JIS X 0203:2012と同様(約9800文字)

(*) ISO/IEC 10646:2003/ Amendment 1:2005の一部として発行された国際標準で、漢字圏で共通に使える日常生活の用を満たす漢字集合

採用
予定

b. 異体字

- 日本語LGRが異体字を定義しない場合、中国語LGRで定義される異体字はRootLGRではすべてブロックされ、どちらの文字も使いたいという要求には応えられない。それを回避し異体字文字列も登録可能とするためには、少なくとも日本語LGRでも保護的(受身的)に中国語LGRが定義する異体字を異体字として定義することは必須である【ケーススタディ参照】
- 考えられる選択肢

- ① 中国語LGRで定義される異体字を日本語LGRでも異体字として保護的(受身的)に定義する
- ② 保護的(受身的)定義に加え、日本語LGRでも独自(積極的)に異体字を定義する
ドメイン名というコンテキストに適切な異体字を、適切な根拠に基づいて定義しオーソリティを得ることは困難である

保護的(受身的)定義によりどの文字列も登録可能となり、日本語JPドメイン名と差異は小さく利用者の混乱は少ないと考えられる。保護的(受身的)で十分かを示すため、日本語における既存の異体字定義候補を定性的・定量的に分析しておくことは必要である

c. 処置

- 考えられる選択肢

① 日本語として登録可能なすべての異体字は登録可能とする

保護的(受身的)異体字定義をする場合は、この選択肢しかない

② 個々の異体字について処置を決める

独自(積極的)異体字定義をする場合は、個々の異体字について処置を決める必要があるが、適切な根拠に基づいて定義しオーソリティを得ることは困難である

d. ラベル評価ルール

- 考えられる選択肢

① 日本語として不自然な文字の並びを排除する
(例えば長音や繰り返し文字が文字列の先頭に
来ないようにする)ルールを積極的に定義する

② 日本語用のラベル評価ルールは定義しない

採用
予定

Rootゾーンに登録されるTLD文字列は審査が行われるので積極的に定義しなくとも
不自然な文字列は排除されると考えられるため、特に混乱は生じないと考えられる

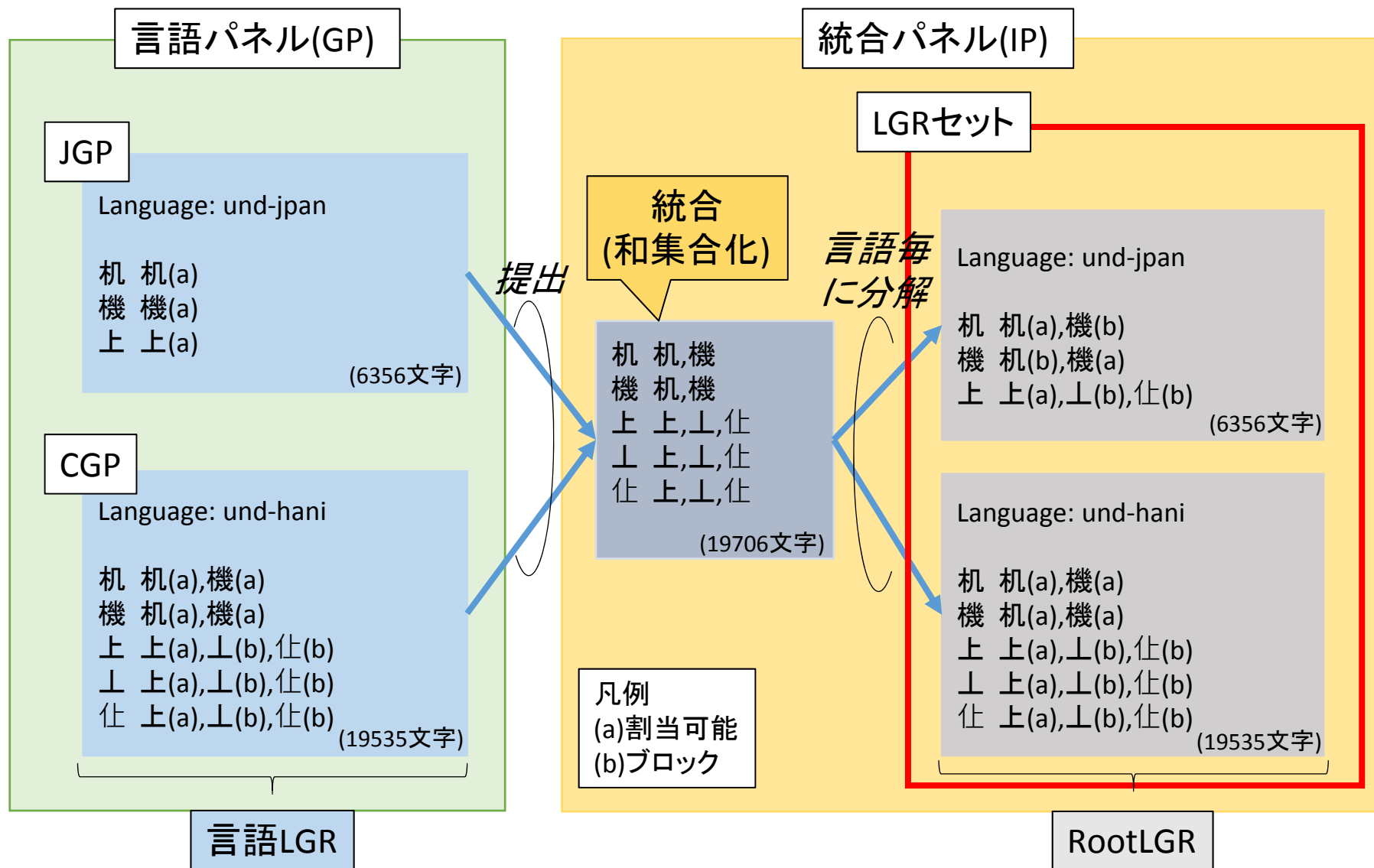
【ケーススタディ】 機上についての考察

- 前提

- 中国語LGRでは「機」と「机」は異体字である
- 中国語LGRでは「上」と「丄」と「仕」は異体字である

(空白)

異体字をまったく定義しない場合



文字列生成例

【注】

混在可能: (a)と(a)

混在不可能: (a)と(b)、(b)と(b)

Language: und-jpan

申請文字列: 机上

割当可能: 机上

ブロック: 机⌊, 機仕, 机上, 机⌊, 机仕

Language: und-hani

申請文字列: 机上

割当可能: 机上, 机⌊

ブロック: 机⌊, 機仕, 机⌊, 机仕

Language: und-jpan

申請文字列: 机⌊

割当可能: 机⌊

ブロック: 机⌊, 机仕, 機上, 機⌊, 機仕

Language: und-hani

申請文字列: 机⌊

割当可能: 机⌊, 機上

ブロック: 机⌊, 机仕, 機⌊, 機仕

Language: und-jpan

申請文字列: 机機

割当可能: 机機

ブロック: 机机, 機机, 機機

Language: und-hani

申請文字列: 机⌊

割当可能: 机上, 機上

ブロック: 机⌊, 机仕, 機⌊, 機仕

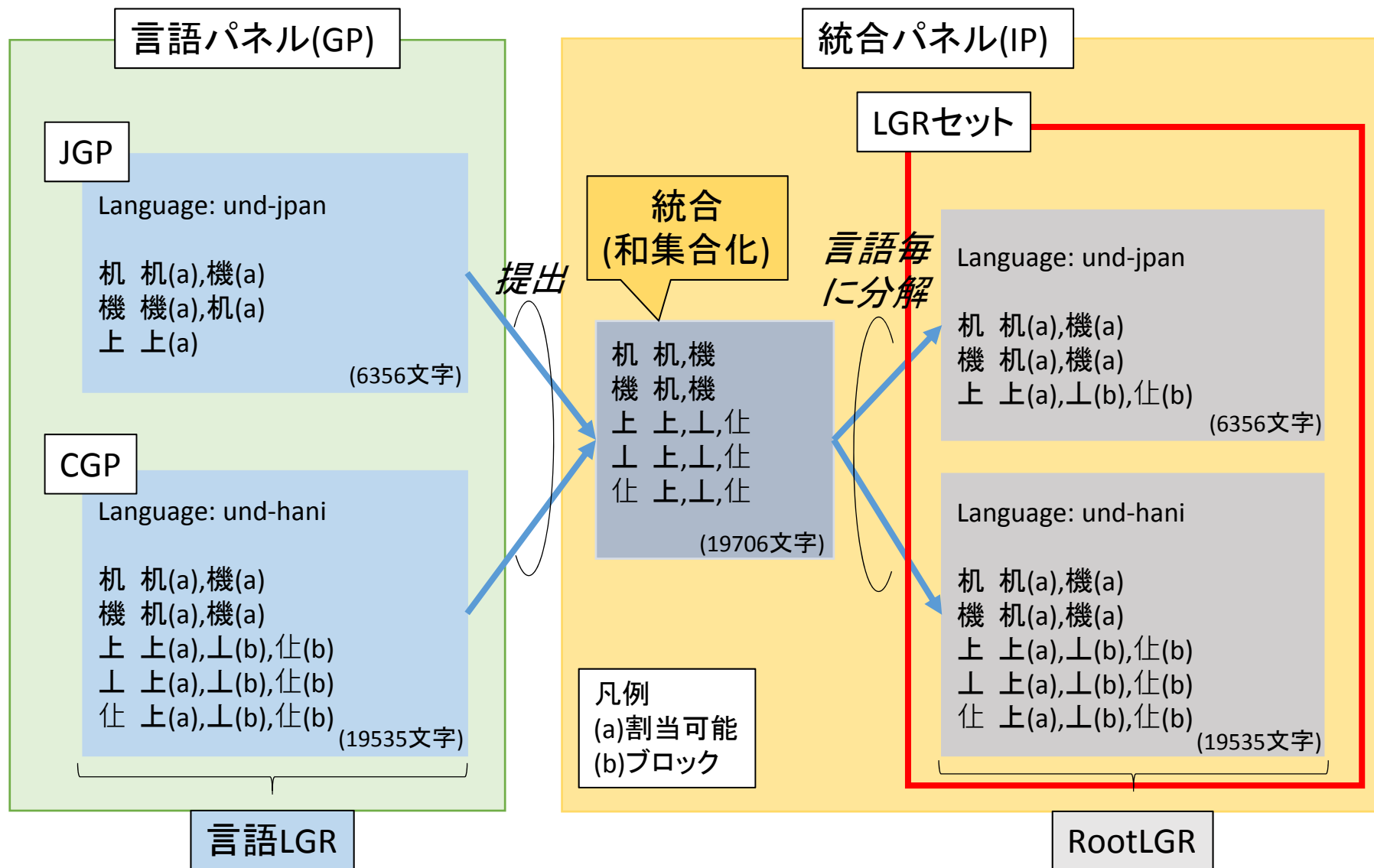
Language: und-hani

申請文字列: 仕仕

割当可能: 上上

ブロック: 上⌊, 上仕, ⌊上, ⌊⌊, ⌊仕, 仕上, 仕⌊, 仕仕

保護的(受身的)に異体字を定義した場合



文字列生成例

【注】

混在可能: (a)と(a)

混在不可能: (a)と(b)、(b)と(b)

Language: und-jpan

申請文字列: 机上

割当可能: 机上, 机上

ブロック: 机上, 机上, 机上, 机上

Language: und-hani

申請文字列: 机上

割当可能: 机上, 机上

ブロック: 机上, 机上, 机上, 机上

Language: und-jpan

申請文字列: 机上

割当可能: 机上, 机上

ブロック: 机上, 机上, 机上, 机上

Language: und-hani

申請文字列: 机上

割当可能: 机上, 机上

ブロック: 机上, 机上, 机上, 机上

Language: und-jpan

申請文字列: 机機

割当可能: 机机, 机機, 機机, 機機

ブロック:

Language: und-hani

申請文字列: 机上

割当可能: 机上, 机上

ブロック: 机上, 机上, 机上, 机上

Language: und-hani

申請文字列: 上上

割当可能: 上上

ブロック: 上上, 上上, 上上, 上上, 上上, 上上, 上上, 上上

【ケーススタディ】 天気についての考察

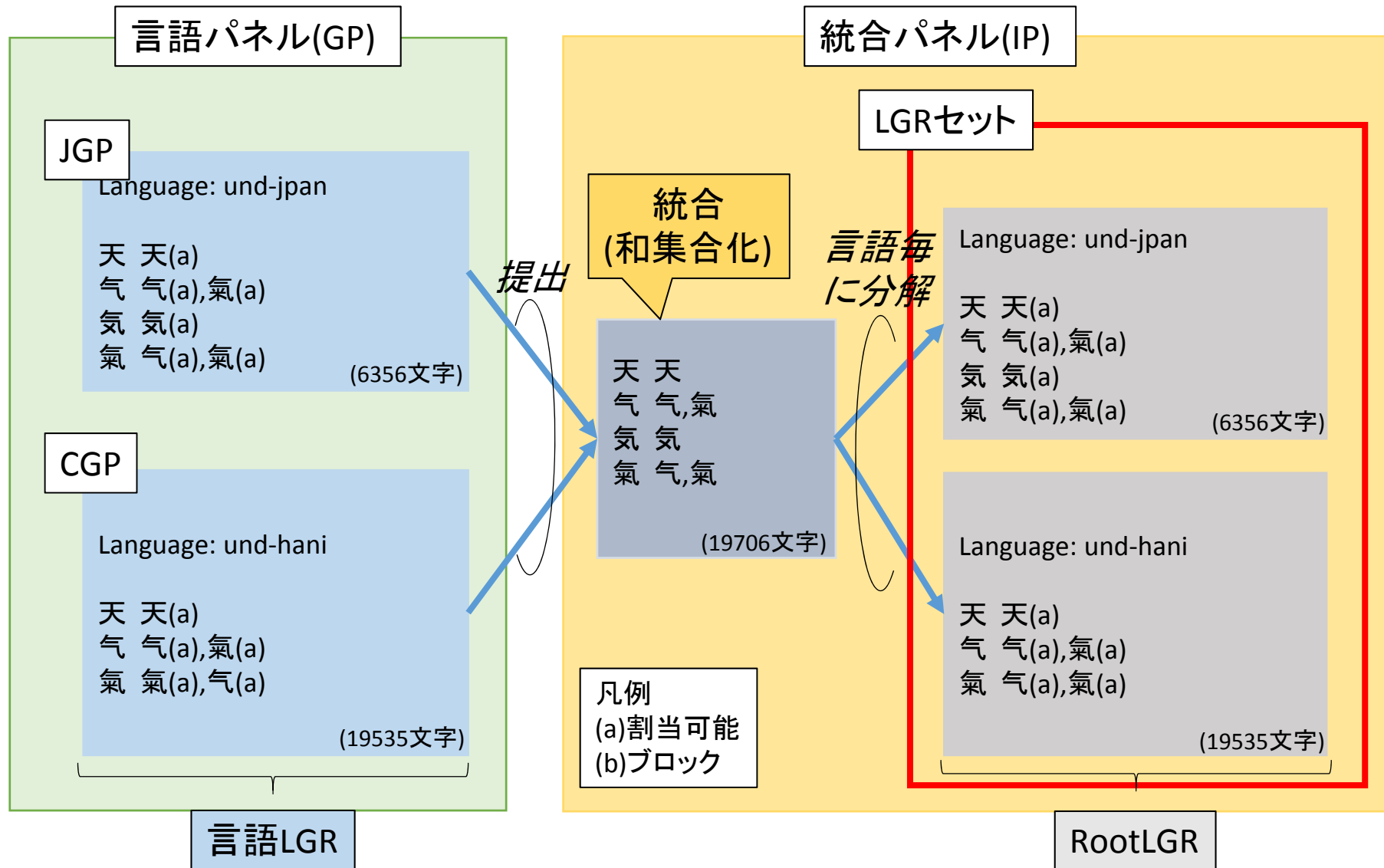
- 前提

- 中国語LGRでは「气」と「氣」は異体字である
- 日本語では「気」と「氣」を異体字とする文献がある
- 「気」は日本語の国字である

ポイント：
天気と天氣をパッケージとする
べきか、独立とするべきか

(空白)

保護的(受身的)に異体字を定義した場合



文字列生成例

【注】

混在可能: (a)と(a)

混在不可能: (a)と(b)、(b)と(b)

Language: und-jpan

申請文字列: 天氣

割当可能: 天氣

ブロック:

Language: und-jpan

申請文字列: 天氣

割当可能: 天氣, 天气

ブロック:

Language: und-jpan

申請文字列: 天气

割当可能: 天气, 天氣

ブロック:

Language: und-hani

申請文字列: 天氣

割当可能: 天气, 天氣

ブロック:

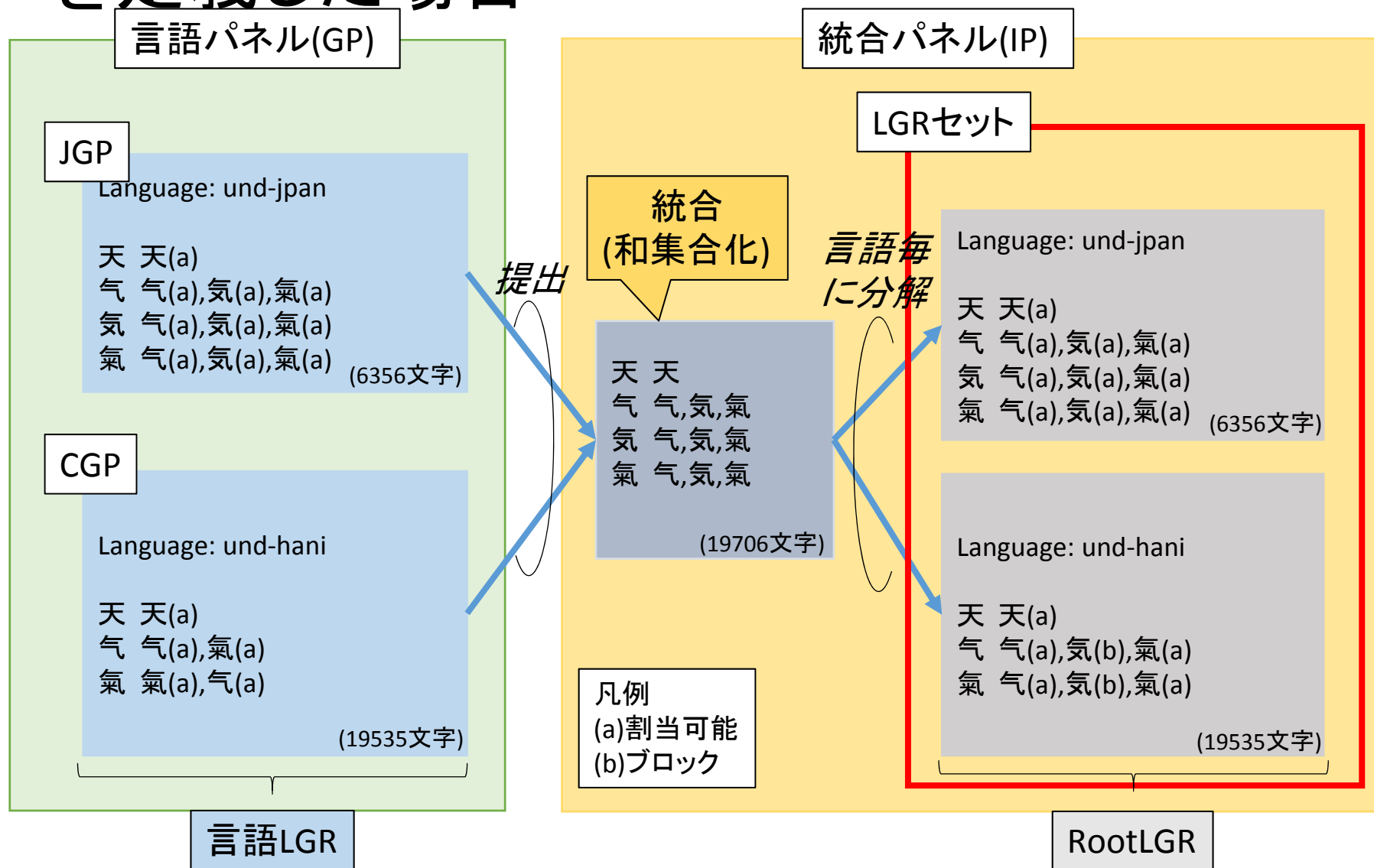
Language: und-hani

申請文字列: 天气

割当可能: 天气, 天氣

ブロック:

保護的(受身的)+独自(積極的)に異体字を定義した場合



文字列生成例

【注】

混在可能: (a)と(a)

混在不可能: (a)と(b)、(b)と(b)

Language: und-jpan
申請文字列: 天氣
割当可能: 天氣, 天氣, 天气
ブロック:

Language: und-jpan
申請文字列: 天氣
割当可能: 天氣, 天氣, 天气
ブロック:

Language: und-jpan
申請文字列: 天气
割当可能: 天气, 天氣, 天氣
ブロック:

Language: und-hani
申請文字列: 天氣
割当可能: 天气, 天氣
ブロック: 天氣

Language: und-hani
申請文字列: 天气
割当可能: 天气, 天氣
ブロック: 天氣