

統合物語生成システムにおける固有名詞概念の体系的記述

A Systematic Description of Proper Noun Concepts in the Integrated Narrative Generation System

寺田貴範[†] 秋元泰介[‡] 小野淳平[‡] 小方孝[†]
 TERADA Takanori AKIMOTO Taisuke ONO Jumpei OGATA Takashi

[†]岩手県立大学ソフトウェア情報学部 [‡]岩手県立大学大学院ソフトウェア情報学研究科

1 まえがき

物語生成に関する従来の考察やシステム(小方・金井, 2010)を集成した「統合物語生成システム(Integrated Narrative Generation System: INGS)」(Akimoto & Ogata, 2012)と呼ぶシステムの構築を筆者らは直近の課題としている(図1)。INGSは処理モジュールとして、物語の世界を設定し時間順の事象構造(ストーリー)を概念レベルで生成する物語内容機構、ストーリーの順序や語り手の視点等の物語言説を操作しその同じく概念構造を構築する物語言説機構、言語・映像・音楽により以上の概念構造を表層表現化する物語表現機構の3つを持つが、これらの処理機構は随所で知識を参照する。図1に見られるように、知識における大きな部分を概念や言語に関する辞書が占める。これまでに開発・利用しているのは、概念辞書(Oishi et al., 2012; Oishi & Ogata, 2012)(名詞概念、動詞概念、形容詞概念、形容動詞概念用)と言語表記辞書(鎌田・小方, 2013)である。既にかなり開発が進んでいる名詞概念辞書が対象としているのは「一般名詞概念」であるが、本稿ではそれとの関連で第一次的な開発を行っている「固有名詞概念」のための辞書について報告する。

辞書である。現状の規模は、動詞概念辞書は意味分類 36 個および動詞概念 11951 個、名詞概念辞書は中間概念 5808 個および終端概念 115765 個である(2014年1月23日現在)。

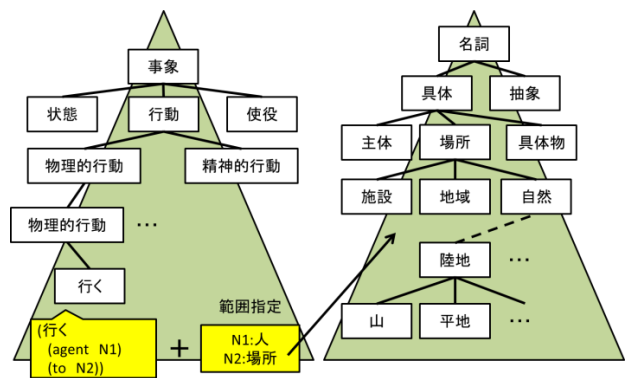


図2 動詞概念辞書と名詞概念辞書の構成概要

本稿で述べる固有名詞概念辞書は、名詞概念辞書における終端概念と関連付けられる。現在はまだ準備されていないが、将来的には固有名詞概念のためにも上述した言語表記辞書を備える予定である。

固有名詞概念がなぜINGSにおいて必要かということについてであるが、これは物語というものの特質と関連する。中には、例えば安部公房の小説のように、故意に物語と関連する固有名詞的な概念を抹消した物語も存在するが、多くの物語では現実存在する具体的な場所・舞台や社会組織等の固有名詞的な概念が出現する。これは個別的なものを通じて普遍に達しようとする物語の一つの戦略である(アリストテレスの『詩学』第9章(アリストテレス, 1997)や『源氏物語』の「蛭」の帖)。また物語における娯楽的性質や広告・宣伝的性質の一つを成すものでもある。例えば大衆的なテレビドラマや推理小説ではしばしば「御当地」の土地や物産等が現れて受け手の好奇心や懐かしさの感覚をそそる。筆者らも INGS のための narrative interface(Ogata & Ono, 2013)の一つの試みの中で岩手県の土地や物産を特別に集めシステムに広告・宣伝的機能を持たせようとした(秋元 他, 2013)。

次に、名詞概念辞書のINGSの処理過程での利用方法を、その記述構成や関連する概念との関わりにおいて説明する。INGSでは、物語内容生成機構において物語内容=ストーリーの構造が生成される。一つのストーリーは、終端要素を事象とし、事象や複数の事象を含む部分構造どうしを結合する関係が中間要素となる。個々の事象は、動詞概念辞書における一つの動詞概念及び名詞概念辞書にお

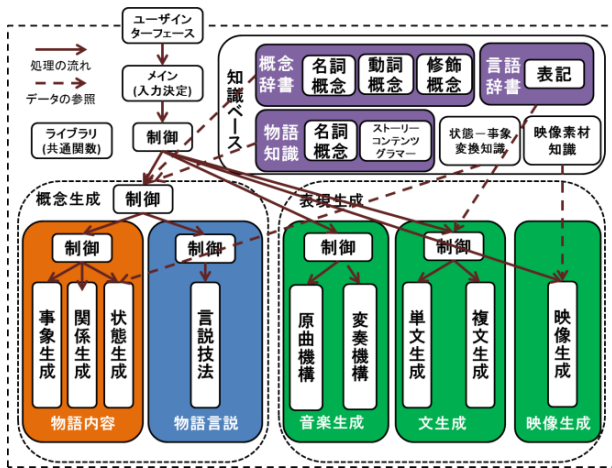


図1 統合物語生成システムの全体構成

2 名詞概念辞書と固有名詞概念

図2にINGSの現状における主要な概念辞書である動詞概念辞書と名詞概念辞書それぞれの全体構成の概要と相互の関係を示す。それぞれ is-a 関係による階層構造を成し、中間的カテゴリー(中間概念)と終端要素(終端概念)を持つ。終端要素が個々の一般動詞概念に相当し、言語表記辞書における該当要素と連携する。言語表記辞書とは個々の動詞概念や名詞概念の文字表示の組を規定するための

る一つ以上の名詞概念と結び付けられた一種の格構造を成す。物語内容生成機構は、以上のような物語内容の構造に対応して、関係構造生成機構と事象生成機構という下位機構に分かれる（さらに状態生成機構もある）。以上のように、名詞概念辞書と関連するのはこのうち事象生成機構の方である。事象生成機構の役割は個々の事象の具体的構造を定義することである。それはまず、対応する動詞概念中に記述された文型パターンとその中の個々の格（主体や目的等）の制約条件により事象の構造を作るが、この制約条件により名詞概念辞書中から可能な名詞概念が選択される。生成された単一の事象の実際の記述例を図3に示す。物語内容生成機構の詳しい機構の説明は(秋元・小方, 2014)を参照されたい。

```
(event 追う2 (type action) (ID 1) (time (time1 time2))
(agent age%看守#1) (counter-agent age%囚人#1)
(location loc%監獄#1) (object nil) (instrument nil) (from nil)
(to nil) (adverb nil) (possessive nil) (situation nil) (purpose nil)
(experiencer nil) (source nil) (idiom nil) (information nil) (as nil))
```

図3 事象の具体的構造の例

ここで、個々の具体的な事象中に格納されている値は「インスタンス」である。名詞概念辞書中の個々の名詞概念はインスタンスに対するクラスであり、名詞概念の一般的な定義である。INGSにおけるインスタンスには幾つかの種類がある。一つは、「特定の名前を持たない最も一般的なインスタンス」である。例えば、教室の中の多数の机の中のある特定の机（例えば登場人物が使用している机）がこのようなものに相当する。次に、「特定の名前を持ったインスタンス」がある。例えば、「太郎」という名前を持ったある登場人物自身やその太郎が飼っている「シロ」という名前の犬がこのようなものに当る。従って、インスタンスと関連付けられるべきこの種の名前の集合を格納しておく必要があるが、これはまだINGSに備わっていない。三つ目が、ここで論じている「固有名詞としてのインスタンス」である。それも名前を持ったインスタンスの一種には違いないが、その名前は一般的なものではなく唯一無二のものである。例えば、単に太郎やシロと言っても対象を一義的に特定することはできないが、岩手山や北上川と言えば特定することができる。（しかし同じ名前が複数の対象を指示する例もある。例えば、「横浜」という土地は下北半島にもある。「川端康成」は他にもいるに違いないが、川端康成と言えば普通あの人を指すだろう。このような意味では固有名詞としてのインスタンスと一般的な名前を持ったインスタンスとの違いは程度問題ということになる。次節で述べるように現在地理的情報に絞って固有名詞概念の体系化を行っているが、実際、「赤男山（あかおとこやま）」、「赤男山（あかおやま）」のような同じ漢字で記述される固有名詞概念も数多く存在する。）このように、固有名詞概念とは、インスタンスの中で一種の公共に属するものを意味する。

このように、物語内容生成機構における事象生成機構によって実際に生成された事象を構成する要素としてのインスタンスは幾つかのタイプに分かれるが、何れも「属性フレーム」に具体的な値が格納された属性集合として記述される。属性フレームとは、一般名詞概念、名前の付いた名詞概念、固有名詞概念のそれぞれの情報内容を表現する

ための属性情報を格納するためのリスト（フレーム）形式としての枠組みである。現状のINGSではその属性の種類を体系的に検討・確定する段階には至っておらず、最低限の属性を設定しているに過ぎない。この属性フレームに基づく、現状での一般名詞概念のインスタンスが格納された属性集合の例を図4に示す。

人物(Agent)	場所(Location)
((ID age%看守#1) (instance-of 看守@番人) (type nil) (location loc%監獄#1) (名前 太郎) (所持 obj%鍵#1) (健康状態 100) (外装 nil) (体勢 屈) (意識 覚醒))	((ID loc%監獄#1) (instance-of 監獄@司法機関[監獄]) (名前 nil))
	物(Object)
	((ID obj%鍵#1) (instance-of 鍵@作業具[開閉][鍵]) (location loc%監獄#1) (名前 nil) (形状 nil) (耐久性 100))

図4 属性フレームに基づく属性集合の具体例

事象生成機構が個別の事象構造を生成する処理の大きな部分は、それを構成する個々の名詞概念と関連付けられた属性フレームを枠組みとして、その値をインスタンス情報で埋めることである。現状では、一般名詞概念や名前の付いた名詞概念の場合、対象名詞概念が含まれる名詞概念辞書中のカテゴリーに基づいてインスタンスのタイプを「人物」、「物」、「場所」の何れかに設定し、タイプごとに共通に定義された属性フレーム（属性の種類と各属性のデフォルト値を定義したもの）を適用する、という方法で行っている。固有名詞概念の場合は、個々の固有名詞概念との関連において記述された属性フレームに基づいて、具体的な値を格納する処理として実現する予定であるが、本稿ではまだそこまでは達しない。本稿で行うのは、一般の名詞概念辞書における終端要素と固有名詞概念とを対応付ける処理までである。

3 名詞概念辞書に対応した固有名詞概念の体系的整理

『日本語語彙大系』(池原 他, 1999)の「固有名詞意味属性体系」では、固有名詞は、「地名」「人名」「組織名」「その他」という大きなカテゴリーに分けられている。そのうち「地名」には「地域名」「自然名」「施設名」が属し、それぞれがさらに下位カテゴリーに分類され、具体的な固有名詞の項目に接続されている。本研究では、手始めとしてこの分類に則って固有名詞概念の分類と一般名詞概念辞書との連携を試みる。今回対象とするのは上記「自然名」の下の「地形名」である（「自然名」にはもう一つ「天体名」が含まれる）。「地形名」の下にはさらに「陸上地形名」「河川湖沼名」「海洋名」「海底地形名」の下に 23379 個の固有名詞が収集されている。これは、「～山」「～川」「～海」等の地理的な固有名詞の名称を含むリストである。しかし、例えば「○○崎」「△△島」「××岬」等が一つのカテゴリー中に区別なく存在している。ここでは、名詞概念辞書における終端概念に対応するようにしながら、図5に示すようにこれらの固有名詞概念を分類することを目標とする。さらに次の作業としてそれらに属性フレームを帰属させる予定であるため、本研究では固有名詞概念という名称を用いている。さらに、これをベースに、固有名詞概念体系としての拡張と精緻化を図る予定である。

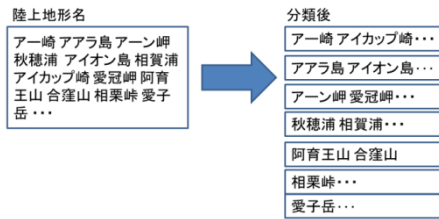


図5 固有名詞概念の分類目標

ユーザが指定した文字と同じ語尾を持つ固有名詞を上記の固有名詞の集合から選択するプログラムにより、共通の語尾を持つ固有名詞のリストを作成した。これと手作業による補助的作業を組み合わせ、23379個の固有名詞すべての分類を行った。この作業を行う際、最初は語尾一文字だけで行ったが、そうすると例えば、「島」と「半島」が同じグループに分類されてしまう。島は「周囲が水によって囲まれた小陸地」、半島は「陸地が海に長く突き出した所」で常識的にも明らかに異なる概念なので、そういった場合には語尾を二文字、三文字と増やして再分類を行った。これらの結果、23379個の固有名詞を216グループに分類した。

次に、これらのグループのそれぞれを名詞概念辞書中の特定の終端概念と対応付けた。現状のプログラムにおいては、(下位概念名@中間概念名(分類した固有名詞のリスト))のように記述している。名詞概念辞書における個々の一般名詞の実際のデータ記述は図6のようにになっている。名詞概念における“terminal” (下位概念)として、分類の際に用いた固有名詞の語尾と共通の単語が存在すれば、その名詞概念の下に対応する固有名詞のリストを関連付けた。しかし分類に利用した語尾が名詞概念における下位概念に存在しない場合は、類似する下位概念を探しそこを関連付けた。さらに、ある下位概念の記述に対応する固有名詞のリストが複数存在する場合は、最も常識的に近いと開発者が判断した名詞概念との関連付けを行った。図6の例では各国の半島が混じり合っている。INGSの構成や目的の観点から改善すべき項目はまだ多い。

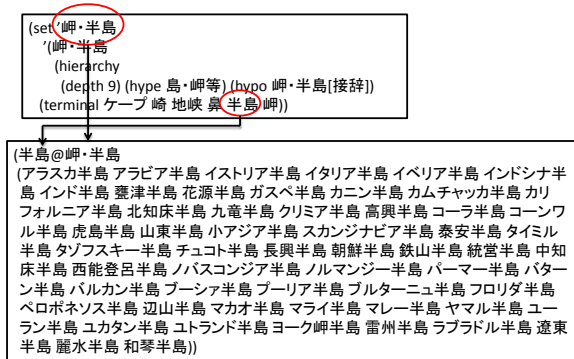


図6 名詞概念辞書と固有名詞の関連付けの例

現状において、名詞概念辞書における94個の名詞概念に、固有名詞概念23379個が関連付けられている。すべての名詞概念と一部の結び付けられた固有名詞概念を表1に示す。残りとして、一般の地理的固有名詞とは思えない総称的な命名と思われるものが含まれており、これは今回の分類からは除外した。また、結び付けるべき名詞概念辞書の終端概念が適当でないと思われる場合もあるので、全

体としてのチェックを行う必要がある。さらに、分類において異なる地域のものが一カテゴリーの中に組織化されずに混在している等の問題もあり、今後は物語生成システムとしての処理の目的に照らし合わせながらの改訂作業も必要である。

4 あとがき—今後の課題—

今後の課題には様々なものがある。以上のような第一次分類の妥当性の確認を経た集成や改訂、分類の精緻化、プログラムとしての実現方式の再検討等が、課題における一つのグループを成す。二つ目は体系化の範囲を地理的情報以外の情報に拡大することである。名前を持ったインスタンスのための人名等の名前の収集と体系化という作業もこの中に含まれる。さらに、重要な課題として、属性フレームの内容の確定と構築がある。この作業は一般名詞概念の属性フレーム構築の作業と一緒にする必要はある。現在、百科事典における項目の最初の概要的情報程度をWikipedia等を素材に用いて自動的に獲得する研究を進めている。収集すべき情報の分量が膨大であるため、今後はこのような物語生成における知識自動獲得やテキストマイニング(小野・小方, 2013)と連動させて研究を進めて行く予定である。

文献

秋元 泰介・今淵 祥平・遠藤 順・小野 淳平・栗澤 康成・鎌田 まみ・小方 孝 (2013). 民話風物語生成・表現システム KOSERUBE 第一版の開発. 人工知能学会論文誌, 28(5), 442-456.

Akimoto, T. & Ogata, T. (2012). Macro Structure and Basic Methods in the Integrated Narrative Generation System by Introducing Narratological Knowledge. Proc. of the 11th IEEE International Conference on Cognitive Informatics & Cognitive Computing, 253-262.

秋元 泰介・小方 孝 (2014). 統合物語生成システムの現状と特に物語内容生成メカニズム. 電子情報通信学会言語理解とコミュニケーション研究会 第4回テキストマイニングシンポジウム予稿集. (印刷中)

アリストテレス (1997). 詩学. アリストテレス詩学, ホラーテュウス詩論. 松本 仁助・岡 道男 訳. 岩波文庫. (第9章).

鎌田 まみ・小方 孝 (2013). 物語生成システムにおける文生成機構と文字表記機構. 言語処理学会第19回年次大会発表論文集, 698-701.

池原 悟・宮崎 正弘・白井 論・横尾 昭男・中岩 浩巳・小倉 健太郎・大山 芳史・林 良彦 (1999). 日本語彙大系. 岩波書店.

小方・金井 明人 (2010). 物語論の情報学序説—物語の思想と技術を巡って—. 学文社.

Ogata, T. & Ono, J. (2013). Designing Narrative Interface with a Function of Narrative Generation. Proc. of the International Conference on Cyberworlds, 214-221.

Oishi, K. & Ogata, T. (2012). The Development of Conceptual Dictionary for Narrative Generation System The Structure and Functions. Proc. of 4th IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 168-170.

Oishi, K., Kurisawa, Y., Fukuda, I., Akimoto, T. & Ogata, T. (2012). Building Conceptual Dictionary for Providing Common Knowledge in the Integrated Narrative Generation System. Proc. of 34th Annual Conference of the Cognitive Science Society, 2126-2131.

小野 淳平・小方 孝 (2013). 物語生成システムにおける表現要素のWebからの自動獲得. 人工知能学会全国大会 (第27回) 予稿集, 214-4in.

表 1 名詞概念と固有名詞概念の一覧（部分）

名詞概念	固有名詞概念	名詞概念	固有名詞概念
流@河川	揖斐川支流 (計 1 個)	尾根@中腹	赤仁尾 穴手尾 池ノ尾 (計 272 個)
畔@田	洞爺湖畔 (計 1 個)	背@中腹	馬ノ背 寒ノ脊 (計 2 個)
田@田	鰻田 河戸新田 (計 2 個)	土手@堤	千間土手 (計 1 個)
海岸@海岸	相馬海岸 宗谷海岸 プッド海岸 (計 18 個)	堤@堤	石江堤 石堤 角助堤 (計 13 個)
浦@海岸	秋穂浦 相賀浦 青砂ヶ浦 (計 662 個)	堰@堤	荒堰 五所川原堰 廻堰 (計 3 個)
磯@海岸	赤磯 阿高磯 荒磯 (計 41 個)	港@公共機関[港]	奈古港 宮崎港 奈古港 (計 3 個)
洋@外海	インド洋 遠洋 鹿島洋 (計 32 個)	津@港	駒津 (計 1 個)
海盆@外海	アルゼンチン海盆 アンガラ海盆 アンゴラ海盆 (計 9 個)	池@池	秋ヶ館溜池 朝日池 芦ヶ池 (計 194 個)
海溝@外海	アカブルコ海溝 アタカマ海溝 アリシユジャン海溝 (計 25 個)	高地@土地	アメリカ高地 ギアナ高地 西北高地 (計 6 個)
沖@外海	川口沖 天神沖 鼻面沖 (計 4 個)	低地@土地	ツラン低地 勇払低地 (計 2 個)
浜@岸	相ノ浜 粟生浜 赤子浜 (計 185 個)	地@土地	北公有地 吉蘭泰塩地 国有埋立地 (計 4 個)
岩@岩石	愛鷹岩 アナマ岩 アフ岩 (計 363 個)	料地@土地	小平蕊御料地 西御料地 (計 2 個)
穴@穴	鬼穴 香木穴 亀穴 (計 8 個)	国有地@土地	西沢国有地 東沢国有地 (計 2 個)
湖@湖	アイセル湖 青木湖 阿寒湖 (計 199 個)	谷地@土地	赤井谷地 宇留井谷地 蒲谷地 (計 24 個)
潟@湖	相内潟 阿世潟 尼潟 (計 37 個)	列島@島	格列飛列島 金門列島 五島列島 (計 17 個)
地方@地域[範囲]	低地地方 (計 1 個)	離島@島	赤離島 内離島 外離島 (計 10 個)
峰@山頂	合ノ峰 相ノ峰 青根ヶ峰 (計 341 個)	本島@島	沖繩本島 本島(ほんしま) 本島(ほんじま) (計 3 個)
頂上@山頂	飛騨頂上 (計 1 個)	島@島	アアラ島 アイオン島 アイランド島 (計 2029 個)
峠@峠	相栗峠 相名峠 アウサリ峠 (計 1435 個)	中島@島	上社中島 毛根中島 野案中島 (計 7 個)
氷棚@山{本体}	氷棚 フィルフナー氷棚 ラーセン氷棚 (計 5 個)	新島@島	大新島 本新島 (計 2 個)
台地@山{本体}	ウストワルト台地 根釧台地 (計 2 個)	孤島@島	沖ノ孤島 地ノ孤島 (計 2 個)
台地@山{本体}	安居台 秋吉台 朝日台 (計 95 個)	群島@島	奄美群島 鞍馬群島 永楽群島 (計 21 個)
山脈@山{本体}	天城山脈 アルタイ山脈 アルチン山脈 (計 63 個)	洞窟@洞穴	青の洞窟 アツツラ洞窟 ラワー(計 3 個)
山地@山{本体}	秦巴山地 御坂山地 養老山地 (計 10 個)	洞@洞穴	悪洞 穴洞 飯田洞 (計 55 個)
山@山{本体}	阿育王山 合窪山 相沢山 (計 6153 個)	窟@洞穴	孔雀ヶ窟 青竜窟 達谷ノ窟 (計 5 個)
高台@山{本体}	上常呂高台 築別高台 西高台 (計 5 個)	大陸@陸地	アジア大陸 アフリカ大陸 アメリカ大陸 (計 13 個)
天地@社会[世界]	天地 (計 1 個)	湾@内海	泉湾 生浦湾 大浦湾 (計 426 個)
林@森林	大倉山林 大矢官林 乙ガ林 (計 17 個)	海峡@内海	明石海峡 アレヌイハハ海峡 岐崎海峡 (計 365 個)
森@森林	青ノ木森 相森 赤倉森 (計 331 個)	瀬戸@内海	芋子瀬戸 牛窓瀬戸 本瀬戸 (計 84 個)
国有林@森林	青井岳国有林 青屋国有林 梓川国有林 (計 93 個)	砂漠@平野	アタカマ砂漠 アラビア大砂漠 インド砂漠 (計 29 個)
水路@水路	荒川放水路 江戸川放水路 西天竜水路 (計 5 個)	平原@平野	華北平原 湖広平原 シベリア平原 (計 14 個)
瀬@州	相寅瀬 相瀬 赤穂瀬 (計 316 個)	野@平野	相野野 饗場野 阿賀野 (計 392 個)
州@州	江南三角州 珠江三角州 (計 2 個)	野原@平野	阿野原 市野野原 井野原 (計 28 個)
沼@沼	相野沼 浅内沼 浅畑沼 (計 189 個)	草原@平野	上砥草原 下砥草原 カンボ(計 10 個)
石@石	旭石 雨垂石 出雲石 (計 104 個)	雪原@平野	大和雪原 (計 1 個)
淵@川	亘淵 エムデン海淵 ネロ海淵(計 127 個)	湿原@平野	川南湿原 駒止湿原 (計 2 個)
澗@川	上澗 静閑澗 柳澗(計 3 個)	平地@平野	青平 赤木平 赤沼平 (計 307 個)
川@川	雨鱒川 胆沢川 伊白良川 (計 2566 個)	砂原@平野	絞兵エ砂原 (計 1 個)
河@川	渭河 運遭河 衛河 (計 57 個)	原@平野	青木原 赤土原 赤穂原 (計 757 個)
海@海	星宿海 アラル海 カスピ海 (計 72 個)	盆地@盆地	四川盆地 諏訪盆地 パリ盆地 (計 18 個)
海台@海	海台 ブレーク海台 (計 2 個)	窪地@盆地	相ノ窪 栗ヶ窪 井手窪 (計 38 個)
沢@谷	梅ノ沢 竈ノ沢 鎌谷沢 (計 807 個)	岬@岬・半島	アーン岬 愛冠岬 相泊岬 (計 414 個)
峡谷@谷	大杉峡谷 大峡谷 太魯閣峡谷 (計 3 個)	鼻@岬・半島	相崎ノ鼻 間鼻 赤石ノ鼻 (計 587 個)
溪谷@谷	赤西溪谷 吾妻溪谷 桂川溪谷 (計 6 個)	半島@岬・半島	アラスカ半島 アラビア半島 イストリア半島 (計 61 個)
鞍部@谷	ポアト鞍部 (計 1 個)	崎@岬・半島	阿翁崎 青砂崎 青島崎 (計 910 個)
地溝@谷	ヨルダン地溝帯 ライン地溝帯 (計 2 個)	畑@畑	相畑 赤根畑 浅畑 (計 136 個)
谷@谷	相之谷 青木谷 相合谷 (計 786 個)	滝@滝	三ツ滝 四滝 四度ノ滝 (計 218 個)