

大型辞典改訂と デジタル言語資源

「日本国語大辞典」第三版改訂の現場から

辞書と言語資源

辞書編纂におけるコーパス活用について

デジタル言語資源と英和辞典

- 日本では、長らく英和辞典編集の現場において、「ネイティブ的発想」を担うものとしてコーパスが活用されてきた。
 - 英語の辞書編集の現場では2000年代からさまざまなコーパスが活用され、また独自のコーパス構築なども行われて「corpus-driven」な編集が目標とされてきた。
- 学習言語であるため、
- 内省が効かないという認識がある。
 - 記述の妥当性について、習得の効率面などの評価が可能。

デジタル言語資源と国語辞典

- 国語辞典編集の現場では「用例カード」という形での言語資源の蓄積はおこなわれたが、オープンな形では蓄積されていない。
 - 独自のコーパス構築などを行われず、主に国立国語研究所が構築・整備するコーパスを参照し、作例や語釈執筆に活用するという形で現在に至っている。
- 第一言語（母語）であるがゆえに、
- コーパスが必要がないという認識（誤解）が醸成される。
 - 記述の妥当性についての評価研究も蓄積が少ない。

コーパスからわかること

- 数量が計測できるようになり、頻度や傾向がわかる。
 - その用法・語形・表記・構文が一般的なのか
 - 共起しやすい語はあるか、あるとしたらどんな語か
 - メタ情報を付与することでより高度な検索が可能となり、たとえば年代情報、位相情報を活用できる。
 - その語が使われるようになった時期や時間的变化の過程
 - 分野・スタイル・社会属性等の広がり
- 記述されている国語辞典は少ない。

コーパスの日本語カバー力

- コーパスも辞書も日本語の縮図を示す設計であるはず。
- 現在利用可能なコーパスをもってしても、従来の辞書の記述や辞書編集者の直感のほうが饒舌なこともある。
→言語資源の規模やバリエーションの不足からか。
- 既にあるコーパスを活用していくとともに、まだデジタル言語資源化されていないものを活用できる形にしていくことも期待される。

参考：小木曾ほか（2026）「小型国語辞典と現代語コーパスは互いの語彙をどの程度カバーするか」

デジタル言語資源とAI

- 今後AIの記述によるテキストが増加すると考えられるが、慎重に扱う必要がある（現時点で、言語表現として似た情報の増幅が見られる）。
 - 日本語のデジタル言語資源は、質・量ともにまだまだ発展途上である。
 - 少しでも多くのバリエーションが観察できるよう、さまざまな言語資源を、メタ情報をたくさん付与してデジタル化してほしい。
- 整備のためのAI活用の可能性

デジタル言語資源への寄与

デジタル言語資源の質・量が育っていく。



より豊かな言語表現・正確な用法が検出可能になる。



辞書編集者が検証する。



デジタル言語資源に活用可能な形の「辞書」を提供する。

エコシステムに

第三版改訂の概要

日国の改訂の歴史

大日本国語辞典 1915 修訂版 1939

日本国語大辞典 1972-1973

日本国語大辞典 第二版 2000-2002

日本国語大辞典 第三版 2032 ???

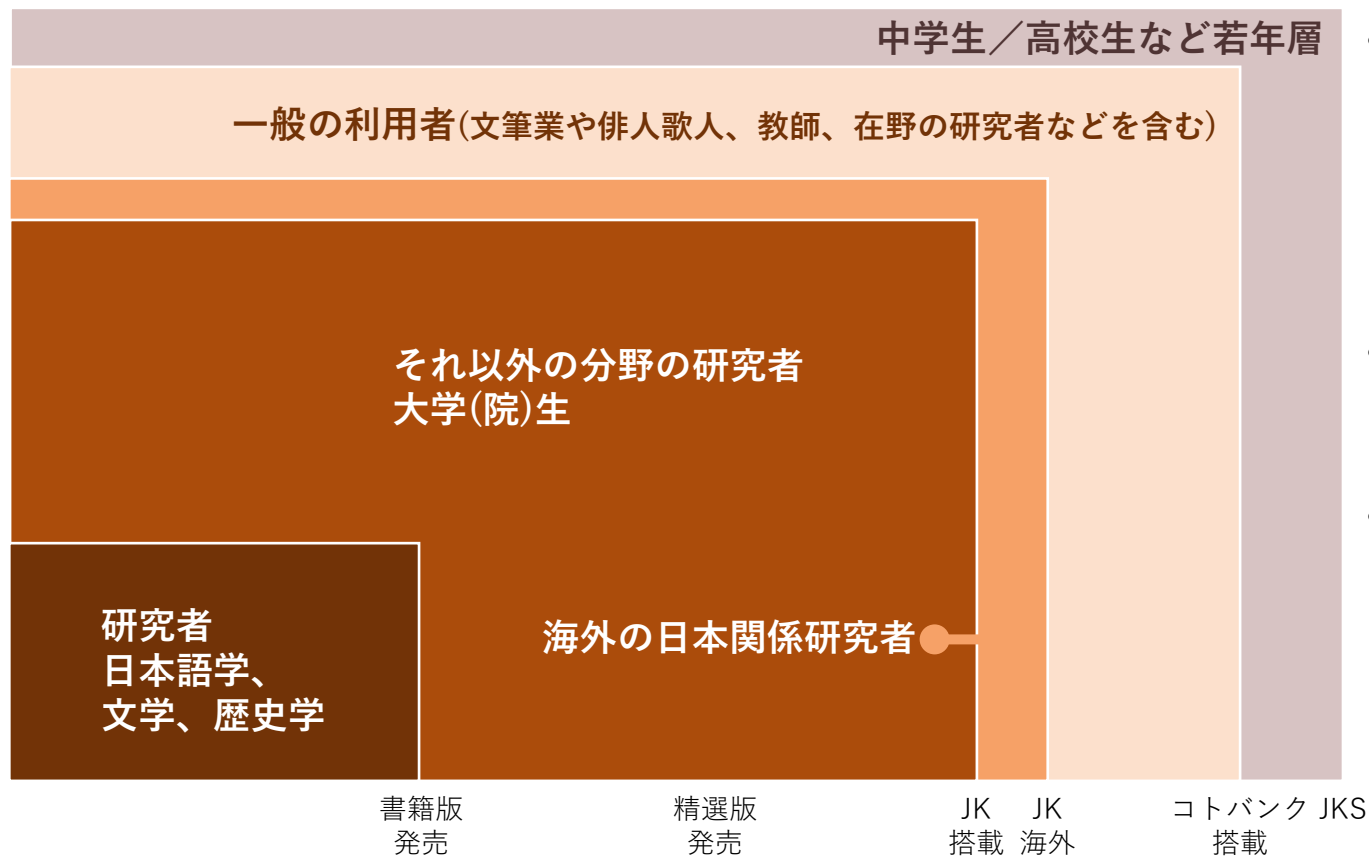
第二版は、日本語や日本文学の専門家だけでなく、各界の識者にも協力をあおぎ、3000人以上が関与した大事業だった。

当時のデータ処理や、組版の最新技術を駆使して制作された。



前身の「大日本国語辞典」以来、20-30年ごとに改訂が行われてきた

増加する利用者



- 第二版刊行後、精選版の刊行や、ジャパンナレッジでの電子版公開、コトバンクでの精選版公開などによって利用者が増えてつづけてきた。
- 現在は、ネットを通じて海外で日本関係の研究をする人々も利用。
- また近年はJKSでの提供によって中学、高校などでも広く利用されている。

あらたな改訂の意義

今の日本語の様相と、約30年間の日本語学の研究を反映させて、日本最大の知的財産を更新する。

- 辞書は実用と研究の接点。定期的に更新（＝改訂）する必要がある。
- 第二版は1980、90年代の情報で編集されており、すでに古い。
- 日本語のインデックスとしても史上空前の規模であり、同等のコンテンツを新たに作ることは困難。編集技術の継承も課題となっている。

日国の特徴

- **用例主義の継承**

幅広い分野から多数の用例を採集。それをもとに立項、説明する。

- **現状唯一の大型国語辞典**

見出し語数50万、用例数100万、全13巻+別巻。

- **古語から現代語まで**

上代から現代まであらゆる文字資料を対象とする。

現代語と古語が、同一の辞書構造の中に併存している。

- **付属要素(各欄)の充実**

補注のほか語誌や古辞書、発音、方言などの要素が豊富。

用例主義の継承

初版以来、膨大な資料から用例を採集し、立項と語釈執筆を行った。これを第三版でも継承。さまざまな言語資源を活用する、新用例主義としたい。

日本国語大辞典

…基本的に実例（実際の用例）を掲出する。

一般的な国語辞典

…基本的には作例を掲出。実例はないか、少数。

- 古典史資料のさらなる検討
- 新しい現代語資料の検討

大量の用例を採集
管理するDBを作成

蓄積された
大量の用例

分析

- 語義の分類（ブランチで分ける）
- 語釈の執筆（意味を説明する）
- 語誌等の執筆

※典型的な例は実際に掲載する

発刊の辞（初版）

国語辞典は一国の文化を象徴する。真の国語辞典の有無、あるいはその辞典の性格に、その国の文化の水準が反映するといつてよい。文化とことばとの深いかわりを考えるとき、一国の文化を継承しこれを将来に伝達するためには果たす国語辞典の役割は、きわめて大きい。（中略）

国語大辞典の生命は、まずその引例文にある。上代から現代に至る実際の用例を集め、その上に立って意義用法を記述すべきである。そのために、国語大辞典の編纂は、さまざまな分野の資料を渉猟して、その中からことばの生きた使用例を集めることから出発する。（後略）

日国＝大型国語辞典

- 今も生きている国語辞典の中で、ただひとつの“大型辞典”。
- 英オックスフォード英語大辞典、仏ロベール仏語大辞典と比較可能。

大型辞典

日本国語大辞典、平凡社大辞典(絶版)、角川古語大辞典(古語)、大漢和辞典(漢字)、日本大百科全書(百科)...

※複数の巻で構成され、何十万項目を収録する辞典。

中型辞典

大辞泉、広辞苑、大辞林、講談社日本語大辞典(絶版)、学研国語大辞典(絶版)、角川大字源(漢字)...

※単巻で刊行され、10万語以上を収録する辞典。机上であつかうことが前提。

小型辞典

新選国語辞典、現代国語例解辞典、例解学習国語辞典、新明解国語辞典、三省堂国語辞典、岩波国語辞典...

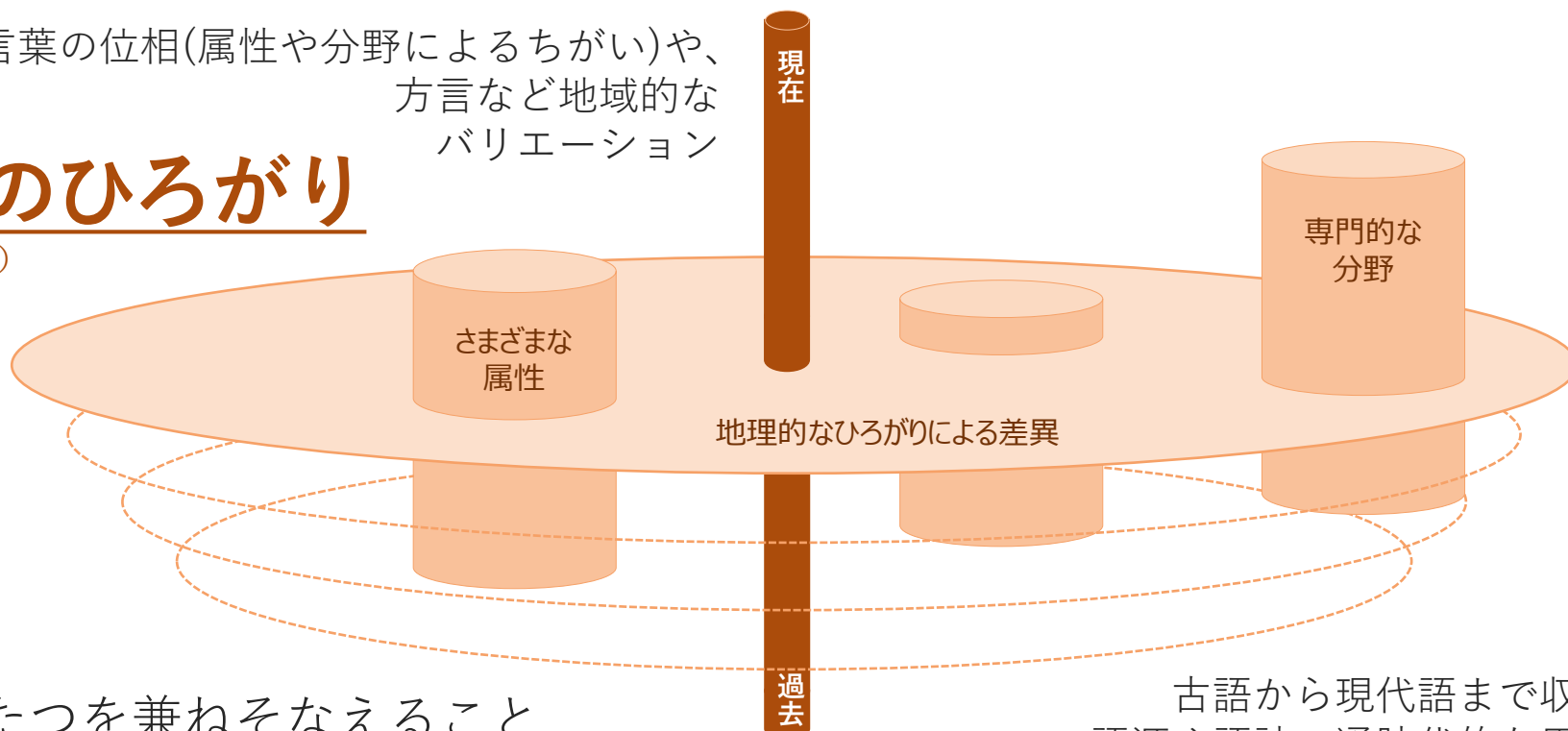
※単巻で刊行され、ハンディであつかいやすく、おおむね10万語までの辞典。

言語の時空間

その言葉の位相(属性や分野によるちがい)や、
方言など地域的な
バリエーション

横のひろがり

(空間軸)



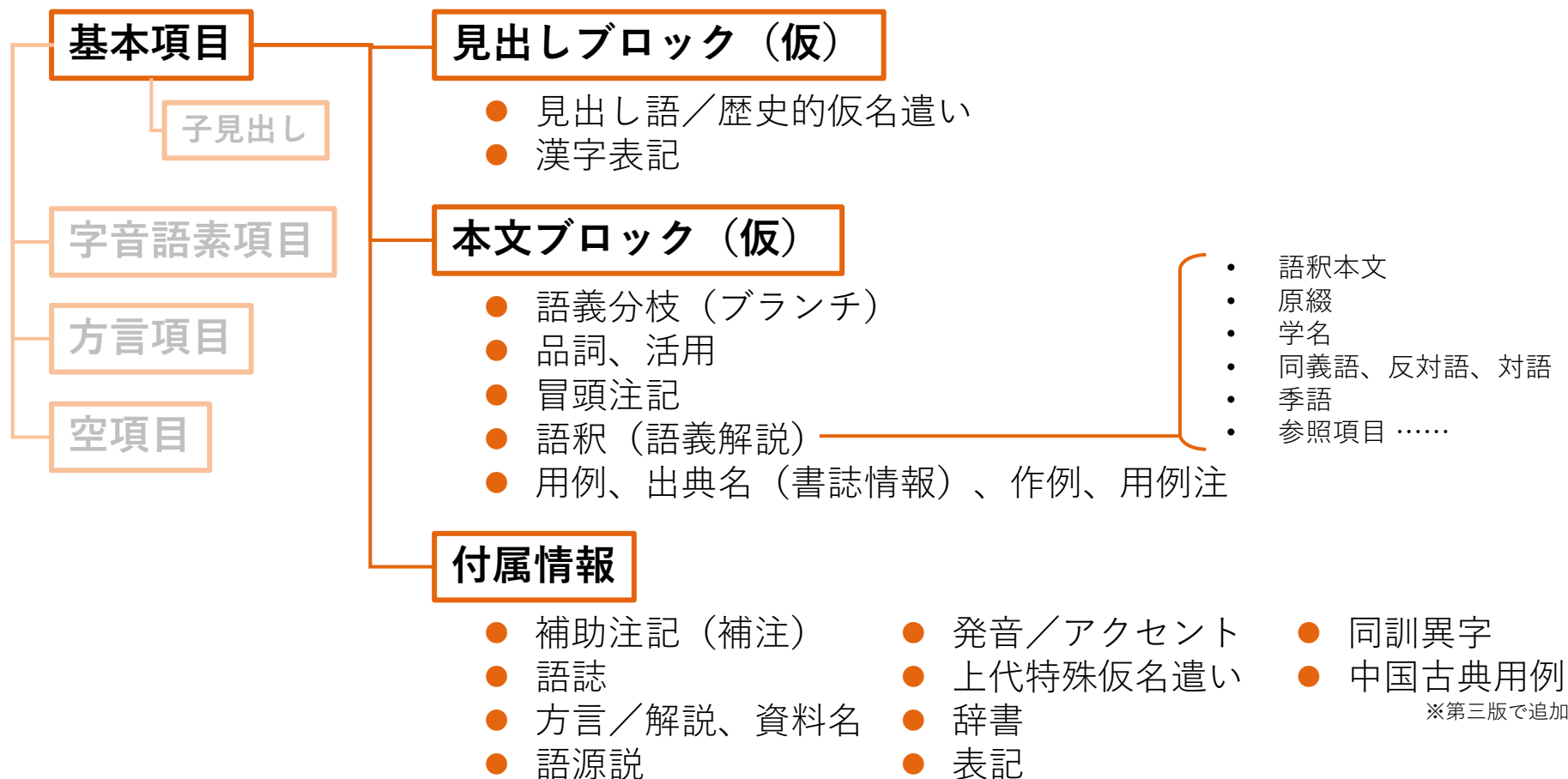
このふたつを兼ねそなえること
で、日本語の全体を俯瞰し、
理解できる唯一の辞典。

縦のながれ

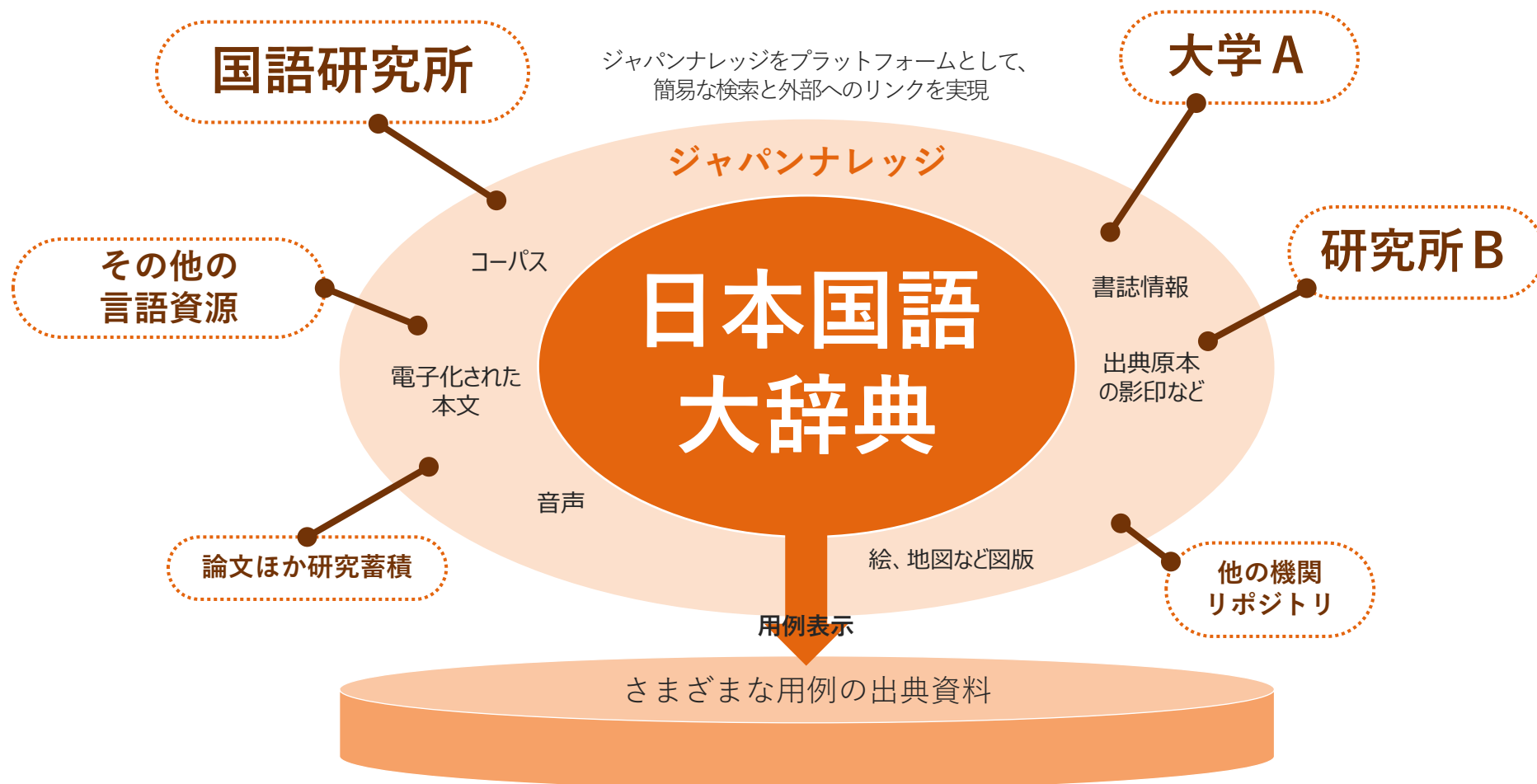
(時間軸)

古語から現代語まで収録。
語源や語誌、通時代的な用例の
列挙など、その言葉の
使い方の歴史も提示

日国の項目モデル



知的情報のハブへ



方言の記述と デジタル言語資源

方言の記述のこれまで

- これまで、各地でそれぞれの方針でつくられた方言資料を収集し、解釈してまとめたものが多かった。
- 共通語と異なる音韻体系を共通語の音韻の範囲におさめたり、共通語と異なる語の単位を共通語の語の単位に合わせたりということがおこなわれている。
- 一定のレベルやルールで整備されたデータはけっして多くないため、活用したくとも難しい。

しかし、急速に各地で「方言」の記録が難しくなっている。

デジタル方言資料の活用

- 低資源言語のAIによる生成など、すでにさまざまな実践も試みられているが、資源のあるうちに記録をできるならしておきたい。
- 語彙、音声や音韻、文法、待遇表現、言語行動そのものの、またそれぞれの時期や時代、年齢別の実態が、一定のレベル・ルールで言語資源化されていけば……。

デジタル方言資料活用例

一定のレベルやルールで言語資源化されていれば、例えば…

- 語彙：共通語・標準語で検索可能
- 音声：辞書の項目・用例と音声を結びつけ、再生可能に
- 音韻：音韻体系の違いを踏まえて／捨象して分析ができる
- 文法：機能語の記述が充実する ex. 「とる」「よる」
- 待遇表現： ex. 敬語に「れる」を多用する地域
- 言語行動：あいさつの地域差

地域差なのか否かが明確になり、日本語の実態がより一層明らかになる。

方言の記述のこれから

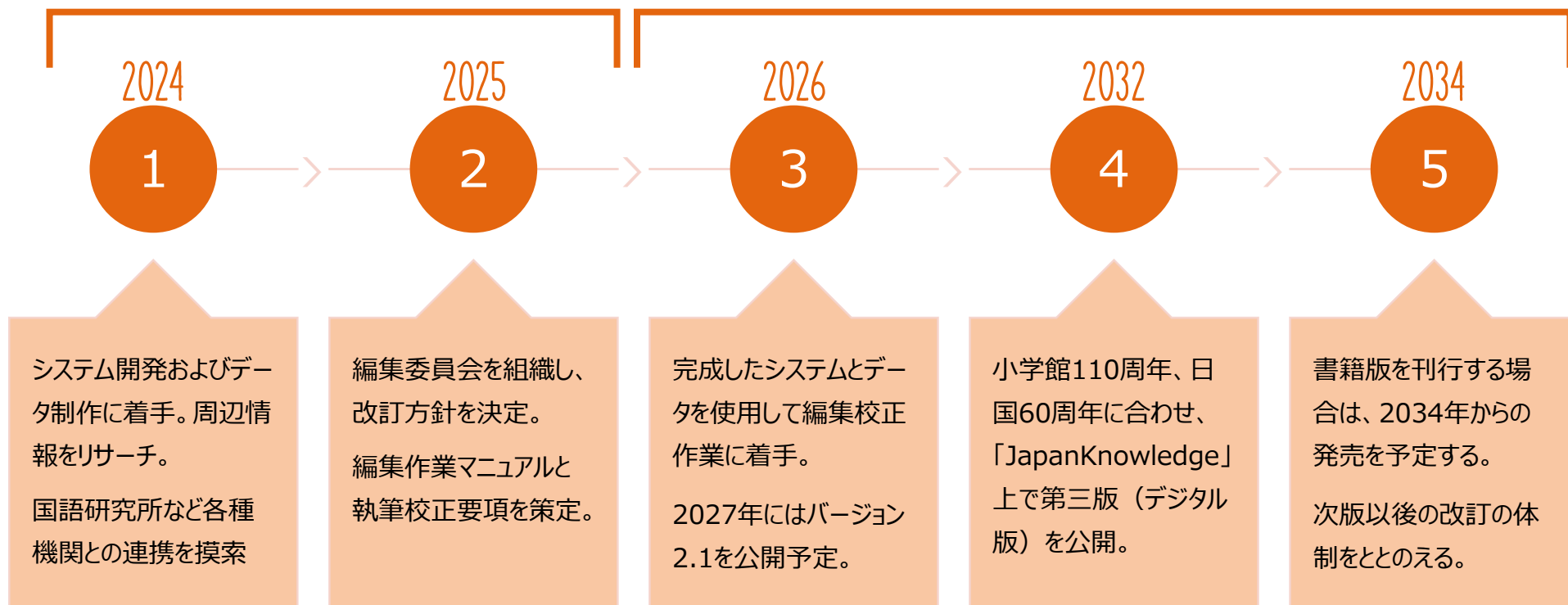
- ある一定のレベルやルールによる言語資源の整備をおこなうことで、より純粹に、地域差について注目しやすくなる。
- 日本語の実態・総体を記述し得る可能性が高まる。
- 「辞書」という活用しやすい形で、方言調査・研究への協力者をはじめとした地域への還元もおこない得る。
- これまで各地で記録されてきたものをデジタル言語資源化することで、古語と方言の関係なども見えてくる可能性もある。

「日本国語大辞典」
のこれから

マイルストーン

環境構築(30%)

編集・校正作業(70%)



辞書に刊行はあれど完成はない。
昨日よりも今日、今日よりも明日。
ひとつでも多くの用例を見つけ出し、
情報を付け加えたい。

松井栄一

2014

「次の三版がどのような形でいつの出版になるかはわからない。
しかし、日国の歴史はたゆみなく続いていく」