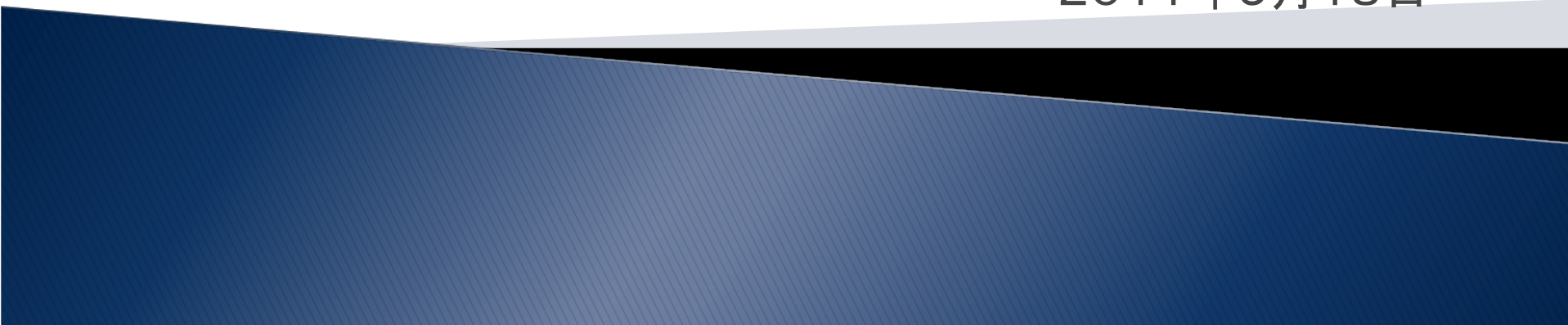


通時辞書の構築について

Kerri L Russell
ケリー・ラッセル

広島大学 special lecture
2011年5月18日



目次

- ▶ オックスフォード大学で行われる日本語史の研究プロジェクト (VSARPJ) の紹介
- ▶ 通時辞書を構築する問題
 - 時代により発音が変わった語
 - 時代により品詞が変わった語
 - 単純語や複合語
- ▶ VSARPJ の通時コーパスで使用する **lexicon** という辞書の紹介

VSARPJ の紹介

- ▶ 「日本語史における動詞の意味構造と項の具現化：日本語史におけるシンタクスの基礎に関する包括的研究」 Verb semantics and argument realization in pre-modern Japanese: A comprehensive study of the basic syntax of pre-modern Japanese
- ▶ 基金：Arts and Humanities Research Council, UK （英国芸術・人文リサーチカウンスル）
- ▶ 期間：2009年1月から2013年10月まで
- ▶ 人事：
 - プロジェクトリーダー Frellesvig (Principal Investigator), Sells (Co-Investigator)
 - フルタイム PD 研究員二人 Horn, Russell
 - 大学院生 1 人
 - その他、十人以上の日本・アメリカ・イギリスの研究者のメンバー
- ▶ url: <http://vsarpj.orinst.ox.ac.uk>

VSARPJ の紹介

▶ コーパスのテキスト

古事記歌	712 年
日本書紀歌謡	720 年
風土記歌謡	730-740 年
仏足石歌	753 年
万葉集	759 年
続日本紀歌謡	797 年
上宮聖徳法王帝説	不明

VSARPJ の紹介

▶ 文法的アノテーション

○ 原文の表記法

- 表意文字
- 表音文字

○ 品詞

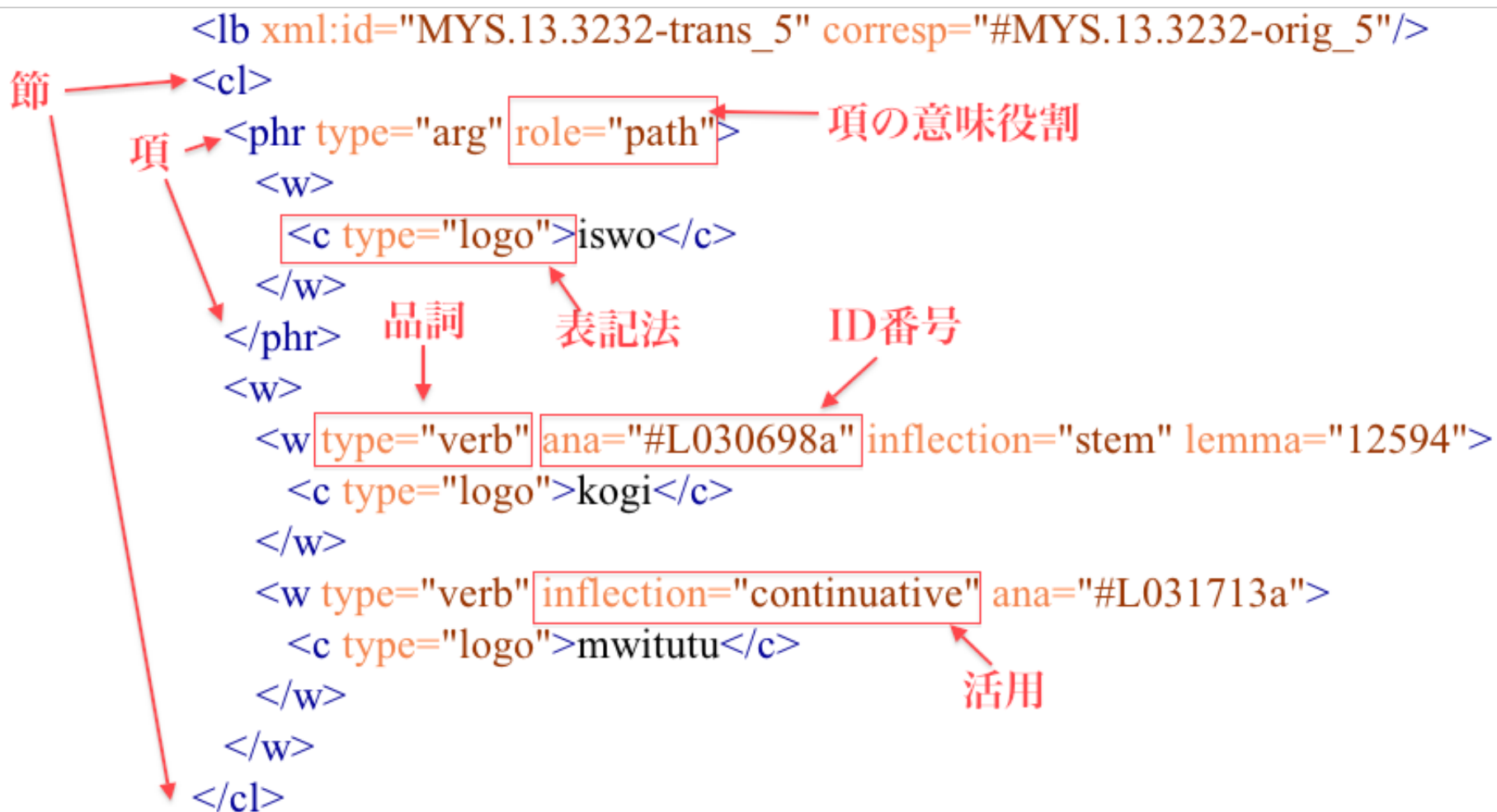
○ 形態素解析

- 接頭語、接尾語
- 活用

○ 語彙素と形態素の ID 番号

○ シンタクス

- 文、節、項
- 項の意味役割



VSARPJ の紹介

- ▶ 現在はオックスフォード大学と国語研は研究協力することになった。
- ▶ オックスフォード大学は上代語コーパスを作成中。
 - 項の意味役割のタグ付け作業中。
 - <http://vsarpj.orinst.ox.ac.uk/corpus/>
- ▶ 国語研の古典コーパスを使って、シンタクスのアノテーションを含む。

通時辞書を構築する問題

▶時代により発音が変わった語

▶母音

- 奈良時代には母音は8つあった。
- 平安時代からは5つあった。
 - かはのへ 「川の辺」
 - 奈良時代：kapanope の「へ」は乙類の「へ」であった。しかし、「辺」は甲類の「へ」だったので、kapanope の意味は「川の上」であった。
 - 平安時代：甲乙の区別しない。意味は「川の辺」に変わった。

通時辞書を構築する問題

- ▶ 時代により発音が変わった語
- ▶ 母音
 - 奈良時代には母音は8つあった。
 - 平安時代からは5つあった。
 - ・ 「問ふ」と「妻問ふ」
 - ・ 表意文字で書かれた時は「問」を使った例はある。
 - ・ 表音文字の場合は「問ふ」は twopu（甲類の「と」）で、「妻問ふ」 tumadopu（乙類の「と」）であった。だから、別語である。

通時辞書を構築する問題

▶時代により発音が変わった語

○子音

- 奈良時代の「p」は平安時代から母音間の「w」になって、頭語の「p」は鎌倉時代ごろ「f」になった。

- 恋：

• 上代語	kwopwi	こ乙ひ乙
• 800 から	kwopi	こ乙ひ
• 950 以前から	kopi	こひ
• 950-1000 ごろから	kowi	こひ
• 1100 ごろから	koi	こい

通時辞書を構築する問題

▶時代により品詞などが変わった語

○如し

- 現代語の観点から見れば、「如し」は助動詞。
- 上代語では形容詞であった。
 - 「ごと」、「ごとし」、「ごとく」などの形容詞の活用で発生した。

通時辞書を構築する問題

▶時代により品詞などが変わった語

○如し

例 1 : (万葉集 11.2352)

玉如	tama no goto
所照公乎	teras-eru kimi wo

例 2 : (万葉集 6.961)

湯原尔	yu no para ni
鳴蘆多頭者	naku asitadu pa
如吾	wa ga goto-ku
妹尔戀哉	imo ni kwopure ya

通時辞書を構築する問題

▶時代により品詞などが変わった語の例

○ながら

- 現代語の観点から見れば、接続助詞である。
- 上代語では接尾語だろう。
 - 名詞
 - kamu-nagara (神ながら)
 - yama-nagara (山ながら)
 - umi-nagara (海ながら)
 - 動詞
 - tuki-nagara (付きながら)
 - tutuke-nagara (続けながら)
 - uresibwi-nagara (嬉しびながら)

通時辞書を構築する問題

▶ 単純語や複合語の問題

- 「顧みる」、「奉る」
- 現代語の観点から見れば単純語である。
- 上代語の観点からは複合語とみるべき可能性もある。
 - 「かえりみる」か「かえる＋みる」
 - 「たてまつる」か「たつ＋まつる」
 - 「たてまだす」という言葉もあった。「挙げる」という意味。
 - 「まつる」という助動詞は奈良時代に使用した。平安時代からは使えなくなった。
 - だから上代語では複合語で、平安時代からは単純語だと考えればよい。
- 更なる研究が必要。

通時辞書を構築する問題

▶ 通時辞書の目標

- VSARPJ の通時コーパスで使用できること。
- 共時の使用を調べられること。
- 通時の使用を調べられること。

lexicon の紹介

- ▶今のところは見出し語が 2514 しかない。
 - 助詞、助動詞、動詞、代名詞、数詞は全部 lexicon にある。
 - 形容詞、副詞、名詞は少ししかない。
- ▶見出しの情報
 - ID 番号
 - 各時代の発音表記
 - 品詞
 - 意味
 - 使用
 - いつから、いつまで
 - 方言
- ▶研究のために、辞書を構築することになった。
 - 研究成果も含む予定。

lexicon の紹介

- ▶ 時代により発音が変わった語の解決法
 - 各時代の発音表記

<superEntry xml:id="L030731-main"> ← 見出し

<entry xml:id="L030731a" corresp="52566">

<form type="stem">

<orth stage="I">kwopwi-</orth>

<orth stage="II">kwopi-</orth>

<orth stage="III">kopi-</orth>

<orth stage="V">kowi-</orth>

<orth stage="VII">koi-</orth>

← 発音表記

<gramGrp>

<pos>verb</pos>

子見出し <iType type="UB"/>

</gramGrp>

</form>

<sense n="1">

<def>love</def>

</sense>

</entry>

<entry xml:id="L030731b" corresp="11800">

<form type="noun">

<orth stage="I">kwopwi</orth>

<orth stage="II">kwopi</orth>

lexicon の紹介

- ▶ 時代により 品詞が変わった語の解決法
 - 使用

```
<superEntry xml:id="L000034-main">
  <entry xml:id="L000034a" corresp="36226">
    <usg only="OJ"/>
    <form>
      <orth stage="I">-mi</orth>
      <gramGrp>
        <pos>adjectival copula</pos>
      </gramGrp>
    </form>
    <sense n="1">
      <def>[infinitive form of adjectival copula]</def>
    </sense>
  </entry>
  <entry xml:id="L000034b" corresp="36226">
    <usg start="EMJ"/>
    <form>
      <orth stage="EMJ">-mi</orth>
      <gramGrp>
        <pos>nominalizer</pos>
      </gramGrp>
    </form>
  </entry>
</superEntry>
```

Diagram illustrating the structure of XML entries for a superEntry:

- The **usg** attribute (only="OJ" or start="EMJ") is highlighted with a red box and labeled "使用" (Usage).
- The **pos** attribute (adjectival copula or nominalizer) is highlighted with a red box and labeled "品詞" (Part of speech).

lexicon の紹介

- ▶ 単純語や複合語の解決法
 - リンクを追加する。

```

<superEntry xml:id="L030517-main">
  <entry xml:id="L030517a" corresp="5975">
    <form type="stem">
      <orth stage="I">kapyerimi-</orth>
      <orth stage="II">kaperimi-</orth>
      <orth stage="V">kawerimi-</orth>
      <orth stage="VII">kaerimi-</orth> }
    <gramGrp>
      <pos>verb</pos>
      <iType type="UM"/>
    </gramGrp>
  </form>
  <sense n="1">
    <def>look back</def>
  </sense>
  <re type="compound" subtype="VV">
    <form>
      <orth>
        <ref target="#L030518a">kapyer-</ref>
      </orth>
      <orth>
        <ref target="#L031712a">mi-</ref>
      </orth>
    </form>
  </re>
</entry>

```

```
<cl>
  <w ana="#L030517a" lemma="5975">
    <w type="verb" inflection="stem" ana="#L030518a" lemma="5979">
      <c type="logo">kapyeri</c>
    </w>
    <w type="verb" inflection="stem" ana="#L031712a" lemma="36920">
      <c type="logo">mi</c>
    </w>
    <m type="auxiliary" function="neg" ana="#L000006" inflection="infconc" lemma="19587">
      <c type="logo">zu</c>
    </m>
  </w>
</cl>
</cl>
```

まとめ

- ▶ 通時的な研究者にとっては、言葉の通時的な使用を調査が容易にできると研究も容易に進む。
- ▶ 更なる研究が必要。

ご清聴ありがとうございました

Kerri L Russell
kerri.russell@orinst.ox.ac.uk

