国立国語研究所学術情報リポジトリ

世界の言語研究所（18）　スロベニアの言語研究所と言語資源　（スロベニア）

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/2165

世界の言語研究所（18）

スロベニアの言語研究所と言語資源（スロベニア）

茂木　俊伸（鳴門教育大学）
アンドレイ・ベケシュ（Andrej Bekeš）（リュブリャナ大学）

1. はじめに

これまでの「世界の言語研究所」シリーズで取り上げられてきた国の中でも，今回のスロベニア共和国(Republika Slovenija)はおそらく最も小さな，日本人にとってあまり馴染みのない国であると思われる。そこで，まずこの国について簡単に紹介しておきたい。

スロベニアは中欧に位置し，西はイタリア，北はオーストリア，東はハンガリー，南はクロアチアと国境を接している。面積は約2万 km^2（日本の四国の面積に相当），人口約200万人の緑豊かな小国である。かつては旧ユーゴスラビア連邦共和国の先進地域であったが，1991年に独立を宣言し，2004年5月には欧州連合(EU)に正式加盟を果たしている。

首都リュブリャナ(Ljubljana)はスロベニアの文化・経済の中心地であり，今回取り上げる研究・教育機関も，この都市に置かれている。公用語であるスロベニア語は，スラブ語派に属しており，話者の数は250万人ほどと言われる（金指 2001）。

さて，ここまでに挙げた数字は，日本(語)のそれと比べると，かなり小ぢんまりとした印象を受けるかもしれない。しかし，興味深いことに，スロベニアでは，未だ日本では実現していない，多ジャンルのテキストを含んだ1億語を超える規模のコーパスが，複数作られている。

大規模コーパスの構築は，その性質上，公的機関が中心的役割を果たすことが多いが，この点はスロベニアでも同様である。以下では，このようなコーパスの整備を行っている二つの研究機関を中心に，スロベニアの言語資源の状況について，簡単に紹介していきたい。

2. フラン・ラモウシュ・スロベニア語研究所

フラン・ラモウシュ・スロベニア語研究所（Inštitut za slovenski jezik Frana Ramovša）[1]は，スロベニア科学・芸術アカデミー(Slovenska akademija znanosti in umetnosti)の付属機関の一つであり，スロベニア語の言語資料や言語資源の作成，および研究活動を行っている機関である（現在の名称は，著名な言語学者である初代所長に由来する）。

その研究の対象は，スロベニア語の正書法，歴史，方言，文法，語彙，音韻など広範にわたっており，多くの辞書・語彙集のほか，モノグラフや紀要の形で成果を公表している。中でも約11万語を収録した『標準スロベニア語辞典』（Slovar slovenskega knjižnega jezika, SSKJ）は，1970年から1991年までの約20年をかけて刊行された辞書であり，合本や逆引き版もある。

同研究所は，オンラインでも言語資源を提供しているが，国内最大のコーパスである『Nova beseda（新「ことば」）』コーパス[2]もその一つである。これは，約1億6,000万語（以下，データはいずれも2005年7月現在）を含む現代語のコーパスであり，バランストコーパスではないが，書籍や新聞，雑誌，書き起こされた話し言葉（国会議事録）といった複数のジャンルのテキストによって構成されている。より詳細なジャンルは，次のとおりである。

フィクション	1,200万語
ノンフィクション	200万語
科学・技術系出版物	200万語
新聞（1998年～現在）	1億2,000万語
国会議事録（1996～2004年）	2,000万語
雑誌	600万語

表1：Nova beseda コーパスの構成

このうち，フィクションには，スロベニアの代表的な文学者であるイワン・ツァンカル(Ivan Cankar)の作品をはじめとするスロベニア語文学と，外国文学の翻訳の両方が含まれている。突出してデータの量が多い新聞は，日刊の『DELO（労働）』紙で，オンラインデータと実際の紙面とを突き合わせながらデータの整理を行っているという。雑誌は，月刊の『Monitor』誌（コンピュータ関連）および『Viva』誌（生活・健康関連）の二種類で，データ量の多い『Monitor』誌の一部は，実際の誌面が参照できる。

Nova beseda コーパスは，オンラインで検索でき，検索結果はKWIC形式で表示される（ただし，著作権上の理由から，キーワードを含む最大3文までの表示に限定されている）。検索の際にジャンルを細かく選択することができるほか，頻度上位1,000語のリストも用意されている。

なお，2005年3月時点の規模と比較すると，ウェブページに掲載されている数字で，フィクション100万語，新聞1,100万語，雑誌200万語の計1,400万語分のデータが追加されており，このコーパスは未だ成長を続けているようである。

3. ヨジェフ・シュテファン研究所

ヨジェフ・シュテファン研究所（Institut Jožef Stefan）[3]は，物理，化学，生化学，電子・情報工学などの分野の研究と大学院教育を行っている機関である。言語研究を主目的とした研究所ではないが，情報工学，自然言語処理のセクションにはコーパス言語学の研究者もおり，さまざまなコーパスやツールの開発が行われている。

同研究所で作られた大規模なスロベニア語コーパスには，『FIDA』コーパス[4]がある。これは，1997～2000年に，リュブリャナ大学文学部，ヨジェフ・シュテファン研究所，出版社DZS，ソフトウェア企業Amebisによって作成された現代語コーパスである（コーパスの名称は，これら4機関・企業の頭文字による）。

規模は約１億語であり，その大半は書き言葉のデータであるが，話し言葉も約200万語含まれている。書き言葉のデータは，書籍，新聞，雑誌といった出版物と，ごく少数の未刊行著作物のテキストから構成される。大まかなジャンル別の構成は次のとおりである。

書き言葉	約１億100万語	（テキスト数29,116）
電子テキスト	約２万語	（同26）
話し言葉	約200万語	（同30）

表２：FIDA コーパスの構成

このコーパスの利用もオンラインで検索する形になっており，検索結果はKWIC形式で表示される。完全版の利用にはライセンスが必要となるが，サンプル版（検索結果を10行まで表示）を無料で試用できる。FIDA コーパスのウェブページの最終更新日は2002年３月であり，これ以降，公開データの更新は行われていないようである。

ヨジェフ・シュテファン研究所が関与している言語資源は，このほかにもある。1995～1997年のプロジェクト MULTEXT-East [5] では，コーパス作成用のタグの規格をまとめた資料類や，タグ付けされた多言語パラレルコーパス（小説『1984年』の，スロベニア語，ブルガリア語，チェコ語といった中欧・東欧諸言語の翻訳版）などが作成された。また，現在も，スロベニア古典資料のデジタル化（翻刻）プロジェクト(Scholarly Digital Editions of Slovenian Literature) [6] などが進行している。

4. 日本語との関わり

ここで，日本語教育との関連から，一つのプロジェクトを紹介しておきたい。

リュブリャナ大学文学部アジア・アフリカ研究学科 [7] の日本研究講座は，スロベニアで唯一の日本語教育・研究機関である。同講座は1995年に設立され，現在では約150人の学生が日本語を学習する，中欧の日本語教育の拠点と言うべき存在になっている。しかし一方で，スロベニア語母語話者用に特化した日本語教材が十分に整備されていないという問題がある。

この問題に対応するために，現在，ヨジェフ・シュテファン研究所の研究者との共同プロジェクトで，オンラインで利用できるスロベニア語・日本語辞書『jaSlo（ヤスロ）』[8] の開発が進んでいる（Erjavec 他 2005）。この辞書は，現在約4,000項目を含む形で公開されており，学習者が，表記や品詞，例文とそのスロベニア語訳などを参照できるようになっている。

5. おわりに――スロベニアの言語資源の背景にあるもの

再び最初に挙げた話題に戻り，小国スロベニアにおいて，上で紹介した大規模なコーパスの整備が何によって後押しされたのか，その要因を少し詳しく考えてみる。これには，歴史的背景とともに，スロベニア語の置かれた現状が大きく関わっている。

もともとスロベニアは，言語を強い拠りどころとして成立した国であり，母語の保護意識は非

常に強い。周知のとおり，イギリスや，スラブ語圏で言えば特にチェコでは，早い時期にナショナルコーパスの構築や研究利用が進められてきたわけであるが，強い言語アイデンティティという土壌を持ったスロベニアでも，これらの海外の動向の大きな影響の下，コーパスの構築や運営に対する使命感を持った研究者によって，二つの研究所を中心に，現在の成果が作り上げられてきた。

加えて，スロベニアを取り巻く言語環境は，EU 加盟によって，旧ユーゴ時代や独立後の10数年間から急速に変化している。EU の現状をスロベニア国内に常に反映させるために，次々に押し寄せる新しい用語，新しい概念を翻訳していく必要が生じ，このこともコーパスに対する実際的な要求をもたらしている（実際，政府の翻訳局では，EU の法令の英語・スロベニア語パラレルコーパス『Evrokorpus』[9]（1,100万語規模）を作成している）。

EU 加盟に伴い，スロベニア語は，英語やドイツ語，フランス語といった影響力の強い言語の中で，まさに生き残りをかけた状況にある。例えば，EU 内の大学に自由に入ることのできる現状では，国内の大学教育のレベルが低下すると，大学生の国外流出を招き，大学で必要とされる高度なスロベニア語も不要になる恐れがある（大学の競争力を維持するために，例えば英語という媒介語を積極的に導入したとしても，同じ結果がもたらされるであろう）。スロベニア語は，小さな言語であるがゆえに，現代の多言語・多文化社会において，どれだけ柔軟で表現力豊かな表現手段として育っていけるかという，切実な問題に直面しているのである。

このような状況の中，スロベニア語コーパスの構築とそれに基づく研究は，言語の実態に合った，正書法や文法の規範，外来語の取り扱いといった言語政策を打ち出す上で，重要な役割を果たすことが期待されている。

注

1 Inštitut za slovenski jezik Frana Ramovša: http://www.zrc-sazu.si/isj/
2 Nova beseda コーパス： http://bos.zrc-sazu.si/a_beseda.html
3 Institut Jožef Stefan: http://www.ijs.si/ijs.html
4 FIDA コーパス： http://www.fida.net/
5 MULTEXT-East: Multilingual Text Tools and Corpora for Central and Eastern European Languages: http://nl.ijs.si/ME/
6 Scholarly Digital Editions of Slovenian Literature: http://nl.ijs.si/e-zrc/
7 リュブリャナ大学文学部アジア・アフリカ研究学科： http://www.ff.uni-lj.si/AzAfr/
8 jaSlo: Japanese-Slovene Learner's Dictionary: http://nl.ijs.si/jaslo/
9 Evrokorpus: http://www.gov.si/evrokor/

参考文献

ジョルジュ・カステラン，アントニア・ベルナール（千田善 訳）(2000)『スロヴェニア（文庫クセジュ827）』白水社
金指久美子（2001）『スロヴェニア語入門』大学書林

重盛千香子（1999）「〔リレー連載：言語ジャーナル53〕スロヴェニア」『言語』28-5，86-87，大修館書店

ロワン，バルバラ（2002）「語のレベルにおける等価性をめぐって－重複部分と相違部分の割合について－」『筑波応用言語学研究』9，59-72，筑波大学大学院博士課程人文社会科学研究科文芸・言語専攻応用言語学領域

Erjavec, Tomaž, Kristina Hmeljak Sangawa, Irena Srdanović, and Anton ml. Vahčič（2005）「日本語・スロヴェニア語ウェブ辞書の開発」『言語処理学会第11回年次大会発表論文集』，703-706，言語処理学会

付　記

筆者らの訪問に貴重な時間を割いて対応してくださった，Primož Jakopin 氏（フラン・ラモウシュ・スロベニア語研究所）と Tomaž Erjavec 氏（ヨジェフ・シュテファン研究所）に御礼申し上げる。また，本稿の作成に際して，Kristina Hmeljak Sangawa 氏（リュブリャナ大学文学部）に重要な助言をいただいた。記して感謝の意を表したい。