

# 神戸市外国語大学 学術情報リポジトリ

## スペイン語における新語抽出

メタデータ	言語: jpn 出版者: 公開日: 2001-10-31 キーワード (Ja): キーワード (En): 作成者: 宮本, 正美, Miyamoto, Masami メールアドレス: 所属:
URL	<a href="https://kobe-cufs.repo.nii.ac.jp/records/950">https://kobe-cufs.repo.nii.ac.jp/records/950</a>

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



# スペイン語における新語抽出

宮 本 正 美

## 1. はじめに

辞書の改訂作業，あるいは，既刊辞書を利用しての新辞書の作成にあたっては，旧辞書には掲載されていない語（これを，「未知語」と呼ぶことにする）を捜し出して必要なものを加えていく作業が必須となる。従来は，国内外の新刊辞書から「未知語」を新見出し語として収集し，その中から必要なものを選別収録するという作業が広く行われていたと思われる。しかし，今日では比較的ローコストで大量のテキストデータを入手することが可能になっている以上，辞書編集者には，既刊辞書を参照するだけでなく，独自にテキストデータから「未知語」を収集する努力が，当然求められる。本稿では，私達が現在採用しているこの「未知語」の抽出法を紹介し，あわせて，「未知語」に含まれる現代スペイン語の「新語」の特徴の一端を記述してみたい。

## 2. 新聞テキストデータからの「未知語」の抽出

ここでは，El País 紙の社会面の記事（1998—2001年度，約42メガバイト，「テキストデータ 1a」と呼ぶ），マドリッドの記事（1999—2001年度，約28メガバイト，「テキストデータ 1b」と呼ぶ）と，ABC, El Mundo 紙など数紙の記事（2001年度，約150メガバイト，「テキストデータ 2」と呼ぶ）をサンプル対象とした。<sup>(1)</sup>

既刊の辞書の見出し語リストをベースに「参照リスト」を作成し，対象とするテキストデータから，「参照リスト」に含まれない語を抽出する。ここ

では、「現代スペイン語辞典」の見出し語に、「西和中辞典」と Vox : *Diccionario general ilustrado de la lengua española* の見出し語の一部などを加筆したものを「見出し語リスト」とする。

「未知語候補」抽出のシェルスクリプトでは、この「見出し語リスト」から、名詞、形容詞などの性数変化を起こす語は、女性形、複数形を派生させ<sup>(2)</sup>、動詞（不定詞）は活用形を派生させる<sup>(3)</sup>。このようにして作成した「参照リスト」は次のような構造になっている<sup>(4)</sup>：

動詞の部分：	非動詞の部分：
(前略)	(前略)
zurcieses	zuta zutas
zurciésemos	zutana zutanas
zurcieseis	zutano zutanos
zurciesen	zutuhil zutihiles
zurciere	zutujil zutujiles
zurcieres	zuzo zuzos
zurciéremos	zuzón zuzones
zurciereis	zwingliana zwinglianas
zurcieren	zwinglianismo zwinglianismos
zurcid	zwingliano zwinglianos

対象テキストの各語形をその頻度数とともにリストアップする（このテキストを「テキスト語形リスト」と呼ぶことにする）。固有名詞など大文字を含む語は今回の「新語抽出」の対象から除くので、「テキスト語形リスト」は、例えば、「テキストデータ 1a」の場合、次のようなものである<sup>(5)</sup>：

488344 de	1 zoosanitaria
261411 la	1 zoroastrismo
204156 que	1 zorra
176729 el	1 zoster
173176 en	1 zozobranter
136768 y	1 zozobró
130505 a	1 zuecos
122045 los	1 zumbando
84101 del	1 zumoo

76496 se

1 zurdo'<sup>(6)</sup>

(中略)

この「テキスト語形リスト」と「参照リスト」を照合すれば、「参照リスト」に含まれない「テキスト語形リスト」の語形、つまり、我々にとっての「未知語候補」が抽出できる。「テキストデータ 1a」の場合、「未知語候補リスト」は次のようなものである：

943 ica	1 yuge
809 ico	1 zanj
621 icos	1 zanjla
458 pa	1 zapeando
361 'vacas	1 zenarol
352 icas	1 zeranol
342 web	1 zetazeta
316 of	1 zoarces
305 com	1 zozobranter
219 pr	1 zurdo'

(中略)

残念ながら、「未知語候補リスト」の最上位中、単語（略語）と言えるものは、458 pa, 361 'vacas, 342 web, 316 of, 305 com だけである。それぞれ、pa (← para), vacas に引用符 ' が付いたもの、英語の web, 英語の of, 略語の .com と考えられる。あるいは、ica, ico などと同様に、対象のテキストの単語が特殊コードを含んでいて、それが正しく変換されていない場合に抽出された文字列だとも考えられる。例えば、従 ica が única と変換されていなければ、ica が文字列として抽出される。以下のように、明らかに単語（略語）とみなせる「未知語」が並べば理想的であるが、現実のコンピュータ処理ではなかなかそうはいかない：

342 web

259 www

89 clonar

48 neuronal

42 internauta

41 módem

36 hackers

34 clon

30 hacker

上記のように、多くのノイズが現れる。いくら前処理を施してもすべてのノイズを排除することは不可能だと考えなければならない。したがって、ここから、つまり、「未知語候補リスト」から我々に関心のある「未知語」を収集するのは手作業になる。

### 3. 現代スペイン語における新語の特徴について

「テキストデータ 1a, 1b, 2」を対象に、以上のような手順で抽出した、頻度数がほぼ5以上の「未知語」から収集した「新語」の特徴のいくつかをここで述べてみたい。

まず第一に指摘しなければならないのは、当然のことながら、「語幹的接辞」<sup>(7)</sup>も含めて接頭辞や接尾辞を付けた派生語が多いということである。接頭辞としては、eco- : ecopunto 「エコポイント (石油, 石炭, 天然ガス, 風力などエネルギー源の環境汚染度)」, euro- : Eurocopa 「(サッカーなどの) ヨーロッパカップ (争奪戦)」, eurozona 「ユーロ (使用) 圏」。inter- : intercentros 「センター間の, 諸機関の間の」, internauta 「ネットサーファー」。macro- : macrorredada 「大型一斉検挙」, macrovertedero 「大規模ごみ廃棄場」。mega- : megabanco 「巨大銀行」, megaproyecto 「巨大計画, メガプロジェクト」。mono- : anticuerpo monoclonal 「単一クローン抗体」。nano- : nanotubo 「ナノチューブ (ナノテク素材, 10億分の1メートル単位の超微細な管)」, nanotubo de carbono 「カーボンナノチューブ」。neo- : neofacha 「ネオファシスト (の)」。pre- : preneandertal 「ネアンデルタール人以前の」, prepago 「前払い, プリペイド」, las tarjetas prepago 「プリペイドカード」。pro- : proetarra 「ETA 支持者 (の)」。re- : realojar 「(...を) 新たな場所に住ませる」, reentrada 「再突入」。sobre- :

sobreexplotar「乱獲する，搾取する」。sub- : subletal「亜致死の，致死量に近い」。super- : superordenador「スーパーコンピューター」，superloto「ジャンボ宝くじ」。tera- : terabyte「テラバイト＝1兆バイト」などが挙げられる。

一方，接尾辞を付けて新しい概念，対象物を表現する新語は，スペイン語で用いられる接尾辞の数を考慮すると，比較的少ない。-ador(a) : pegador(a)「(テニス)ハードヒッター，(ゴルフ)ロングヒッター，(ボクシング)ハードパンチャー」，tuneladora「トンネル掘削機」(← túnel + adora)。-ador(a)は普通，動詞に付加される接尾辞とされているが，これは名詞+adoraである。-al/-ar : corneal「角膜の」，neuronal「ニューロンの」，protocolar「儀礼的な」。それぞれ，córnea+al，neurona+al，protocolo+arから形成される。protocolarの場合は，宮本(1997:75)が指摘するように，語基の最終音節に /l/ が含まれるので，異化作用によって，-alではなく，-arが付加されると考えられる。-aje : patrullaje「パトロール作戦，哨戒行動」(← patrullar + aje)，puntaje「得点」(← punto + aje)。-ante : infartante「感動的な，すばらしい」(← infartar + ante)。-azo : cuerpazo「セクシーな肢体」，pendulazo「激しい振り子運動」(← péndulo + azo)，petardazo「爆発物(によるテロ)」(← petardo + azo)，suedazo「高給」(← sueldo + azo)。cuerpazoとsuedazoの-azoは「評価辞」で「好ましいニュアンス」を加えている。-ero/a : pepero/a「Pepe(例えば，現スペイン首相 José María Aznar)を支持する(人)」，rapero/a「ラップ(ミュージック)の(歌手)」(← rap + ero; 英語 : rapper)。「評価辞」の-ón/ona : derechona「(軽蔑的に)極右，右派(政党)」，muñecón/ona「他人の言いなりになる人」，ofertón「(特別)バーゲン」，poblachón「大きいがごちゃごちゃした町」。-podo : saurópodo「竜脚類(ブラキオサウルス，ディプロドクスに代表される巨大草食恐竜)」。過去分詞を作る接尾辞-ado/aを付加して形容詞，名詞を形成している例として，fugado/a「逃亡した，

脱獄した；逃亡者，脱獄者]，cerebros fugados 「流出頭脳」(≡ fugar + ado)，infartado/a 「(心筋) 梗塞の(患者)」(≡ infartar + ado) をここに加えておこう。

pepero/a のように固有名詞からの派生新語もいくつも見られる。Chechenia の派生語は checheno/a 以外に，頻度は低いが chechenio/a 「チェチェン(人)(の)」が現れる。Felipe から felipismo 「Felipe González (スペイン社会党党首，前首相) の政策，その政治傾向」，felipista 「フェリペ主義(者)(の)」。lepero/a 「Lepe (Huelva 県の村) の(人)，抜目ない(人)」などである。新語とはかならずしも言えないだろうが，lepero のように，村や町の拡大やその話題性とともな地名形容詞もその認知度が高まる。マドリッドの地区や，マドリッド県の村や町をとってみても，例えば，次のようなものが抽出された：-ero/a : alcorconero/a 「Alcorcón の(人)」。-eño/a : algeteño/a 「Algete の(人)」，argandeño/a 「Arganda の(人)」，fuenlabreño/a 「Fuenlabrada の(人)」，parleño/a 「Parla の(人)」，roceño/a 「Las Rozas の(人)」。-ano/a : vallecana/a 「Vallecas の(人)」。固有名詞が普通名詞に転用されているものに，商標の Velcro から普通名詞「マジックテープ」として用いられる velcro や colacao 「コラカオ(ミルクに溶いて飲むチョコレート味の粉末)」，一見英語に見えるが，日本製の外来語でもある cutter 「カッターナイフ」も，Olfa Cutter 「オルファークッター」(≡ 「折る刃」カッター)，NT Cutter 「NT カッター」などの商標に由来する。また，「頭文字語」の GRAPO 「10・1 反ファシズム抵抗グループ」という組織名からそのメンバーを指す grapo/a という語も転用の一つと見なせる。

第二に，英語からの外来語が極めて多い。20世紀後半，特に世紀末からの米国による軍事的経済的な世界支配を反映して，今後もしばらく続く世界的な傾向であろう。政治の世界だけでなく個別言語の世界においても不安定な状況を生み出している。after hours 「早朝まで営業している(バー，ディ

スコなど)], burger 「ハンバーガー店=hamburguesería」, chat 「チャット (ネットワーク上の文字ベースの会話)」, donuts 「ドーナツ」, Grand Slam 「グランドスラム (テニスやゴルフで一人の選手が主要な4大会に優勝すること; 野球の満塁ホームラン)」, hacker 「ハッカー」, hardcore 「ハードコアポルノ (の)」, minivan 「(車) ミニバン」, píxel 「ピクセル (コンピュータなどの画像の最小単位)」, skin head 「スキンヘッド=cabeza rapada」, snowboard 「スノーボード」, streaker 「ストリーカー (ストリーキングをする人)」, web 「ホームページ; ウェブサイト」, さらに, dotcom を西訳したと思われる puntocom 「IT 関連企業」など, 世界に広がる米国文化を象徴している。他の外国語としては, フランス語, イタリア語が少し見られる。18, 19世紀には圧倒的な影響力を誇ったフランス語<sup>(8)</sup>が, 現在では英語にその立場を譲っているのは他の言語においても同様であろう。coupé 「(車) クーペ (の)」, rue 「(口語) 通り」, crutones 「クルトン (細の目に切ったパンをバターや油であげたもの)」などがフランス語から, また, イタリアの監督 Fellini の映画 (1960年) タイトルからくる dolce vita 「甘い生活, 放蕩生活」はもはや新語とは言えないだろう。lehenndakari (lehendakari の綴りが圧倒的に多いが) 「バスク自治州政府首相」に象徴される, バスク語のような地方語も散見される。kale borroka 「街頭闘争」, ¡Gora Euskadi y viva España! 「バスク万歳, スペイン万歳!」の gora 「万歳!」など。

第三に, 複合語, 略語もいくつか見られる。cóctel molotov 「火炎瓶」と同義の bomba molotov, carro lanzaaguas 「放水車」, moto-caca (あるいは motocaca) 「犬の糞回収バイク」, salvapantallas 「(コンピュータ画面の) スクリーンセーバー」, sanecanes 「犬の糞収集容器」, などの複合語。apartahotel (あるいは, aparthotel, apartotel) 「長期滞在用のホテル, ホテル風のサービスのあるアパート」は apartamento+hotel から作られた「省略結合語」(acrónimo)<sup>(9)</sup>である。Kb は kilo byte(s) 「(コンピュータの



処理単位) キロバイト」(kb と小文字で綴ることもある) の、www は World Wide Web 「ホームページ; ウェブサイト」(WWW と大文字で綴られることもある) の「頭文字語」(sigla) である。「頭文字語」が元来何と言う句に由来するかは、ovni 「UFO」が objeto volador no identificado なのか objeto volante no identificado なのかという有名な例を引くまでもなく、時に混乱を生んでいる。先に挙げた GRAPO 「10・1 反ファシズム抵抗グループ」は、Grupos de Resistencia Antifascista Primero de Octubre の「頭文字語」として現れる事例が多いが、Grupos Revolucionarios Antifascistas Primero de Octubre の例も少しある<sup>(10)</sup>。誰もが日常的に用いる IVA 「付加価値税, 消費税」でさえ、Impuesto sobre el Valor Añadido とされていて、この頻度が確かに高いのではあるが、Impuesto del Valor Añadido や、Impuesto al Valor Agregado といった事例までも現れる。

第四に、今回は少なかったが、記事の分野、内容によっては córtex cerebral 「大脳皮質」、cripton, kripton, krypton 「クリプトン(元素の一つ)」、ftalato 「フタル酸塩」、Helicobacter pylori 「ピロリ菌(らせんバクテリア)」(Helicobacter pilory と綴る)、naltrexona 「(麻薬やアルコール) 依存症の治療薬」、nanotubo 「ナノチューブ」、prion 「プリオン(狂牛病の感染原因とされるタンパク質)」、voltarén 「ボルタレン(解熱剤)」といった専門用語、科学用語もたくさん抽出されるはずである。

第五に、これは新聞が対象ということで比較的まれではあるが、俗語、隠語も見られる<sup>(11)</sup>。例えば、chapero 「売春するホモ」(← chapa 「売春婦/夫との性交」+ero)、clefa 「(吸って遊ぶ) シンナー」、julai, julái, julay 「間抜け; あいつ, 奴(軽蔑); おかま」、maketo/a, maqueto/a 「(スペインの他の地域から) バスク地方への移住者(の)」(maketo/a と綴る方が多い) などである。

最後に今回抽出した新語の文法について見ておこう。アクセント位置に各種の見られるものがいくつもあった。「アヤトラー（イスラムシーア派の最高指導者）」の *ayatolá* には *ayatola* もある。「ベルベル人（の）；ベルベル語」の *bereber* には、*beréber* も *berebere* もある。「（ヒンズー教の）導師；（精神的な）指導者」の *gurú* には、*guru* もある。いずれも、外来語というアクセント位置が不安定な語と言える。このように外来語の場合、発音、アクセント位置の表示が必要な語がかなり見られる：*atrezzo/atrétso* / 「（演劇、映画）大道具，小道具」，*burger/búrger* / 「ハンバーガー店」，*coupé/kupé* / 「（車）クーペ（の）」，*disc jockey/disyókei* / 「ディスクジョッキー＝*pinchadiscos*」，*dolce vita/dólche bíta* / 「甘い生活，放蕩生活」，*donuts/dónus* / 「ドーナツ」，*freak/frik* / 「（社会からの）逸脱者（の），ヒッピー（の）」，*minivan/minibán* / 「（車）ミニバン」，*bomba molotov/molotóf* / 「火炎瓶」，*playboy/pléiboi* / 「プレイボーイ」，*skin head/eskín hed* / 「スキンヘッド」，*streaker/estriker* / 「ストリーカー」，*terabyte/terabáit* / 「テラバイト」，*underground/andergráun* / 「アンダーグラウンド（の）」，*yankee/yánki* / 「アメリカ人，アメリカの＝*yanqui, yanki*」など。

新語の文法的性は（特に、外来語の場合）原語が文法的性を持つか否かで扱いが異なる。フランス語，イタリア語のように文法的性のある外来語の場合は，それを踏襲するのが原則と言える：*coupé* 「（車）クーペ（の）」男性，*crutones* 「クルトン（←*croûton*）」男性，*rue* 「通り」女性，*atrezzo*，*atrezo* 「（演劇，映画）大道具，小道具」（←*attrezzo*）」男性，*dolce vita* 「甘い生活，放蕩生活」女性，などは原語通りである。*rue* 以外は，それぞれの語末形「-é，-ón，-o は男性語末形，-a が女性語末形という原則<sup>(12)</sup>」にかなっているからと考えることもできる。一方，英語のように文法的性を持たず，しかもスペイン語にとって特殊な語末形を持つ語の場合にあっては，男性扱いするのが原則である。これは，スペイン語文法においては，「男性が文法的に無標の性である<sup>(13)</sup>」ことによる。「男性が無標」であることを示す文

法事象はいくつも挙げられる：

- a. ellos 「彼ら」; los 「それらを, 彼らを, あなたたちを」; suyos 「彼(ら)の, 彼女(ら)の, あなた(たち)の」, bonitos 「きれいな」など男性複数形の主語/直接目的人称代名詞, (所有)形容詞は, 男性(名詞)ばかりの対象のみならず男性+女性(名詞)の対象も指すのに対して, ellas 「彼女たち」, las 「それらを, 彼女たちを, あなたたちを」, suyas 「彼(ら)の, 彼女(ら)の, あなた(たち)の」, bonitas 「きれいな」など女性複数形は, 女性(名詞)ばかりの対象しか指すことができない。例：Tus ojos (y tu boca) son bonitos/\*bonitas. 「君の目(と君の口)はきれいだ」。
- b. 男性名詞の複数形には, 男性(名詞)ばかりの対象のみならず男性+女性(名詞)の対象も指すものがあるのに対して, 女性複数形は, 女性(名詞)ばかりの対象しか指すことができない。例：los padres 「父親たち, 両親」/las madres 「母親たち」, los reyes 「国王たち, 国王夫妻」/las reinas 「女王たち」。
- c. 不定詞, 副詞, 疑問詞, 間投詞, 句, 文など, 語彙的性のないものは男性扱いされる。例：el vivir 「生きる」, el mañana 「明日」, el porqué 「なぜ」; un pum 「パン, バーン」, el tictac 「チックタック」; gota a gota 「ポタポタと」⇒un gota a gota 「点滴」; el sálvese quien pueda 「我先, 我がち」。
- d. 女性の進出度の低い職種<sup>(14)</sup>の職業名詞では男性形は女性対象も指示するが, その逆は不可。例：el/la médico 「男性/女性医師」; \*el/la médica 「\*男性/女性医師」, el/la jefe 「男性/女性課長」; \*el/la jefa 「\*男性/女性課長」, el/la presidente 「男性/女性議長(大統領)」; \*el/la presidenta 「\*男性/女性議長(大統領)」。
- e. 「動詞+名詞」から成る複合名詞は, 圧倒的に男性名詞<sup>(14)</sup>(あるいは両性名詞)である。例：el/\*la elevalunas 「(自動車の)パワーウインドウ」,

el/\*la salvapantallas 「スクリーンセーバー」, el/la aparcacoches 「駐車係」。

f. 男性形（語基）に接辞が付いて形成される女性形（有標）は多いが、その逆はない。例：profesor 「先生（男性）」⇒ profesora 「先生（女性）」, rey 「王」⇒ reina 「女王」, actor 「俳優」⇒ actriz 「女優」。

g. 男性名詞の方が女性名詞よりも多い。今回使用した「見出し語リスト」<sup>(15)</sup>で、男性名詞、女性名詞、両性名詞の頻度数は：男性名詞（20478 52.18%）、女性名詞（16999 43.32%）、両性名詞（1766 4.50%）。

これらの事象からも「男性は無標」と考えられるので、英語からの外来語で、しかもスペイン語ではまれな語末音（あるいは語末字）で終る名詞は男性扱いされる。例：cyborg 「サイボーグ」, streaking 「ストリーキング」<sup>(16)</sup>, funk 「ファンク（ブルース的要素を加えた黒人ジャズ）（の）」, quarterback 「（アメリカンフットボールの）クォーターバック」, Grand Slam 「グランドスラム」, sex shop 男性（女性）「ポルノショップ」<sup>(17)</sup>, chat 「チャット」, talk show 「トークショー、有名人との会見番組」。この原則にあわない場合は、相当するスペイン語名詞を連想するか否かで「揺れ」が生じていると考えられる。例：jet set 女性/男性「有閑階級」。女性扱いは、clase や gente を連想しているのではないだろうか？ puntocom 女性「IT 関連企業」は empresa（あるいは compañía）を、la London School of Economics 「ロンドン経済学院」は escuela を連想していると思われる。web の女性「ホームページ」、男性「ウェブサイト」は、それぞれ、página と sitio を、after hours 「早朝まで営業しているバー（男性）、ディスコ（女性）」は、それぞれ bar と discoteca を含意している。

複数形についても問題になるのは、ほとんど外来語の場合である。その原語の複数接辞を語末に付けるか無変化の扱いがほとんどである。例：

disc jockey 「ディスクジョッキー」 ⇒ disc jockeys, donuts 「ドーナツ」 ⇒ 無変化, freak 「(社会からの) 逸脱者 (の), ヒッピー (の)」 ⇒ freaks, gay /gai, gei/ 「男性同性愛者 (の)」 ⇒ gays (あるいは無変化), kayak 「カヤック」 ⇒ kayaks, minivan 「(車) ミニバン」 ⇒ minivans, playboy 「プレイボーイ」 ⇒ playboys, playoff 「(スポーツ) プレーオフ」 ⇒ playoffs (あるいは無変化), puntocom 「IT関連企業」 ⇒ 無変化, referéndum 「国民投票」 ⇒ 無変化 (あるいは referendos), skin head 「スキンヘッド」 ⇒ skin heads (あるいは無変化), talk show ⇒ talk shows (あるいは無変化), terabyte 「テラバイト」 ⇒ terabytes, underground 「アンダーグラウンド (の)」 ⇒ 無変化。スペイン語化が進むとスペイン語の複数語尾をとるものが現れる。例: eslogan 「スローガン」 ⇒ eslógenes。

#### 4. 結び

以上、「参照リスト」と「語形リスト」の照合という極めて単純な「未知語」の抽出法について述べた。単純なだけに、対象テキストがいくら大きくても、この抽出法は有効であると考えられる。今日、「新語」を抽出する対象となるテキストが無尽蔵にあることを思えば、今回サンプルとした200メガバイト強のテキストはあまりに小規模ではあるが、その中からもいくつかの傾向が確認された。語彙的、文法的には特に次の4点が挙げられる: 1. 接頭辞、接尾辞の付加による派生新語が多い。2. 英語からの外来新語が多い。3. 英語からの外来新語は、原則的に、「スペイン語において無標の性である男性」扱いされる。4. その複数形は、英語の複数接辞を付けて形成されるか、単数形と同形である。今後の課題としては次のようなものが残る: 変化形の極めて多い「未知」動詞は、個々の頻度数が低い場合、見過ごされてしまう可能性がある。これをどう処理するか。<sup>(18)</sup> また、after hours 「早朝まで営業している (バー/ディスコ)」, bomba molotov 「火炎瓶」, carro lanzaaguas 「放水車」, Helicobacter pylori 「ピロリ菌」などの複合新語を、

連語抽出の立場からどのように効率的に抽出するか、など。<sup>(19)</sup>

注:

- (1) 「テキストデータ 1a, 1b」は白水社編集部の水谷久和氏から、「テキストデータ 2」は小学館マルチメディア局の中村隆宏氏から提供していただきました。いくつかの新語の語義・発音については Francisco Calabuig 氏の協力を得ました。
- (2) 女性形を派生したのは、Vox (1987) から加筆した見出し語の場合である。正確に複数形を派生するためには、単語の分節が必要である。ここで利用した分節スクリプトは、Miyamoto (1997:337-339) をベースにしている。本稿での処理はすべて、AWK (一部 Perl) スクリプトとシェルスクリプトによって Linux 上で行った。
- (3) 動詞活用形の派生手順については、宮本 (1998:20) を参照。
- (4) 行数 約51万6千行, フィールド数 約63万1千, バイト数 約715万9千。
- (5) 行数 約13万4千行, フィールド数 約26万8千, バイト数 約159万7千。
- (6) zurdo に引用符の ' が付いたものと考えられる。
- (7) 高垣 (1995:403-406) 参照。
- (8) 宮本 (1996:131-132, 140-141) 参照。
- (9) 英語では、Beard (1998:57-58) の laser (←lightwave amplification by stimulated emission of radiation), aids (←acquired immune deficiency syndrome) や Joseph (1998:360) の cpu (←central processing unit), ram (←random-access memory) のように、複数の語の語頭の文字を組み合わせて作られる語を acronym と称し、これを大石 (1988:8.1.3) は「頭文字語」と呼んでいる。Beard の smog (←smoke+fog), motel (←motor+hotel) のように、複数の語の一部 (普通、最初の語の前の部分と第二の語の後の部分) を組み合わせて作る語を blending と称し、これを大石は、「混成語」と呼ぶ。一方、Casado (1999:78.2) は前者、即ち、UNED (←Universidad Nacional de Educación a Distancia), pyme (←pequeña y mediana empresa) のように、それぞれの語の最初の文字素 (grafema) を結合したものを sigla と称し、これを宮本 (1996:169-171) は「頭字略号」と呼ぶ。後者、即ち、2語 (まれに3語以上) の1つまたはそれ以上の音節 (最初の語は先頭を、第二の語は語末) をつないで作られる語を Casado (1999:78.3) は acrónimo と称し、これを宮本 (1996:168) は「省略結合語、末端結合語」と呼ぶ。本稿では、前者を「頭文字語」(sigla), 後者を「省略結合語」(acrónimo) と呼ぶことにする。
- (10) 比較的最近の辞書の中で定評のある、Maldonado (1996) と Gutiérrez (1996) は Grupos de Resistencia Antifascista Primero de Octubre を、一方、Seco, Andrés y Ramos (1999) は Grupos Revolucionarios Antifascistas Primero de Octubre を挙げている。実際に記者がマニュアルとして参照しているとは思えないが、El Mundo 紙, El País 紙の Libro de estilo も GRAPO を Grupos de Resistencia Antifascista Primero de Octubre の略号としている。
- (11) 比較的最近の隠語、俗語の意味の変遷については、Casado, Garrido y Miyamoto (1981:83-86) 参照。

- (12) スペイン語名詞の文法的性と語末形の数量的な相関関係については, Miyamoto (1995:74-81) を参照。
- (13) 男性が無標であるという指摘は, Academia (1973:179), Bosque (1983:137), Ambadiang (1999:74. 2.2.6) などにも見られる。
- (14) Val Álvaro (1999:73.3.5) も, 複合名詞に関する文献では, 「動詞 + 名詞」から成る複合名詞が圧倒的に男性だと繰り返し指摘されていると述べている。
- (15) 厳密には, 「見出し語リスト」から, 一部加筆した Vox (1987) の見出し語を除いたリストを対象にした。ここでは, 男性形と(もしある場合には)それから派生した(あるいは対応する)女性形が並置されている。
- (16) cyborg, streaking のように, -g で終る語は /θibor/ (あるいは /sáibor/), /strikin/ (あるいは /estrikin/) のように /g/ を落して発音されることが多く, その場合に現れる /t/, /n/ はそれぞれ圧倒的に男性名詞に多い語末音である。名詞の性と語末文字については, 宮本 (1995:24-26) などを参照。
- (17) Seco, Andrés y Ramos (1999) は, 男性あるいは女性として, 女性扱いの例を1つ挙げている。Maldonado González (1996) と Gutiérrez Cuadrado (1996) は男性としている。我々のコーパスでも6例すべて男性扱いであった。
- (18) 「既知」動詞の変化形の頻度数を集約する方法は, 宮本 (1998:20-21) 参照。
- (19) 連語の一般的な抽出法については, 宮本 (1998) 参照。「狂牛病」が, la enfermedad de las vacas locas と呼ばれたり, あるいは el mal de las vacas locas と呼ばれたりするように, 新しい連語は語彙的にも興味深い。

### 参考文献

- Abad, Francisco y Antonio García Berrio (1983): Introducción a la lingüística, Alhambra, 1983.
- Academia Española, Real (1973): Esbozo de una nueva gramática de la lengua española, Espasa-Calpe, 1973.
- Ambadiang, Théophile (1999): "La flexión nominal. Género y número", Bosque y Demonte (1999), pp.4843-4913.
- Beard, Robert (1998): "Derivation", in Spencer & Zwicky (ed.) (1998), pp.44-65.
- Bosque, Ignacio (1983): "La morfología", Abad y García Berrio (1983), pp.115-153.
- Bosque, Ignacio y Violeta Demonte (1999): Gramática descriptiva de la lengua española, 3 vols., Espasa-Calpe, 1999.
- Casado Velarde, Manuel, Miguel Angel Garrido Gallardo y Masami Miyamoto (1981): "Slang Used by Present-day Spanish Young People", 関西外国語大学研究論集, Vol.34, pp.83-98, 1981.
- Casado Velarde, Manuel (1999): "Otros procesos morfológicos: acortamientos, formación de siglas y acrónimos", Bosque y Demonte (1999), pp.5075-5096.
- El Mundo, libro de estilo, Ediciones Temas de Hoy, 1996.
- Gutiérrez Cuadrado, Juan (1996): Diccionario Salamanca de la lengua española, Santillana, 1996.

Joseph, Brian D. (1998): "Diachronic Morphology", in Spencer & Zwicky (ed.) (1998), pp.351-373.

桑名一博 他編 (1990): 「西和中辞典」小学館, 1990.

Libro de estilo, El País, Ediciones El País, 1996.

Maldonado González, Concepción (1996): *Clave, diccionario de uso del español actual*, Ediciones SM, 1996.

Miyamoto, Masami (1995): "Ensayo para un análisis de la estructura silábica, acentual y morfológica en el léxico español", *Lingüística Hispánica*, Vol.18, pp.61-92, 1995.

——— (1997): "Sobre la estructura del léxico en *Cien años de soledad*", Torre y García Barrientos (1997), pp.329-340.

宮本正美 (1996): "近代および現代のスペイン語", 山田 他 (1996), pp.125-188.

——— (1997): "コンピュータ スペイン語学のすすめ", *NHK テレビ スペイン語会話*, 7月号, pp.72-75, NHK出版, 1997.

——— (1998): "El Mundo 紙における連語の自動抽出", *神戸外大論叢*, Vol.49, No. 2, pp.3-27, 1998.

大石強 (1988): 「形態論」, 開拓社, 1988.

Seco, Manuel, Olimpia Andrés y Gabino Ramos (1999): *Diccionario del español actual*, 2 vols., Grupo Santillana de Ediciones, 1999.

Spencer, Andrew and Arnold M. Zwicky (ed.) (1998): *The Handbook of Morphology*, Blackwell, 1998.

高垣敏博 (1995): "語形成", 山田 他 (1995), pp.389-414.

Torre, Estéban y José Luis García Barrientos (1997): *Comentarios de textos literarios hispánicos*, Editorial Síntesis, 1997.

山田善郎 他編 (1990): 「現代スペイン語辞典」白水社, 1990 (初版), 1999 (改訂版).

山田善郎 他 (1995): 「中級スペイン文法」白水社, 1995.

——— (1996): 「スペインの言語」同朋舎出版, 1996.

Val Álvaro, José Francisco (1999): "La composición", Bosque y Demonte (1999), pp.4757-4841.

Vox (1987): *Diccionario general ilustrado de la lengua española*, Bibliograf, 1987.