

スペイン語コーパスによる受動文の検索

宮本 正美・高垣 敏博

はじめに

近年わが国においても、語学研究者がコンピュータとデジタル・コーパスを用いて、言語研究を行う試みがますます盛んになってきている¹。本稿は、コンピュータ言語学 (Computer Assisted Linguistics) とコーパス言語学 (Corpus Linguistics) の立場から、現代スペイン語コーパスに基づいてなされたスペイン語の受動文に関する高垣・宮本 (2003) や高垣 (2003), 高垣 (2004), Takagaki (to appear) などの研究で用いられた分析手法について述べることを目的とする。

なお、テキスト処理はすべて Linux 上で、AWK あるいは Perl によるスクリプトとシェルスクリプト²を使用して行った。

1. スペイン語の受動文

まず、スペイン語受動文をコーパスによって研究する目的と問題点を略述しておこう。

1.1 SER 受動文と SE 受動文

スペイン語には2種類の受動文が存在する。

(1) a. La noticia fue divulgada por los periódicos. 〈迂言的受動文=SER 受動文〉

その知らせは新聞によって広められた。

b. Aquí no se venden periódicos. 〈再帰受動文=SE 受動文〉

ここでは新聞は販売されていない。

(1a) のようないわゆる迂言的受動文 (以下, 「SER 受動文」) は英語の受動文と同じように, コピュラ動詞 *ser* と他動詞の過去分詞を用い, 他動詞目的語を主語とし, 前置詞 *por* により導かれる動作主句を随意的にとることができる。

一方, スペイン語には(1b)のように再帰構文を用いて「事物」のみを主語と

する再帰受動文（以下、「SE 受動文」）も見られるが、ふつう *por* 句によって動作主を明示することができない。したがって受動文主語が「人間」の場合は SER 受動文しか用いることができないが、主語が「事物」であり、かつ *por* 句を用いて動作主を明示する必要がないときには SER 受動文と SE 受動文とが競合することになる。

ところで、一般に、スペイン語の SER 受動文は英語などの場合と違って使用頻度がきわめて低いことがよく指摘される。反対に SE 受動文の使用度がこれを大きく上回っていることも容易に直感できるところである。またこのような指摘は研究や文法書などにも言及が見られる³。

1.2 コーパスによる受動文の生産性調査

De Miguel (1992:199-218)によると上述のような SER 受動文の低い生産性にはいくつか理由が考えられるが、そのひとつは、SE 受動文が SER 受動文に代替し、その領域を侵食していることがあげられる。もう一つの要因としては、助動詞であるコピュラ動詞 *ser* のもつ時制アスペクトと過去分詞となっているもとの動詞の語彙アスペクトの制約が考えられる。英語の受動文に比べると、スペイン語の SER 受動文は動詞が完了(perfective)時制であるのか未完了(imperfective)時制であるのか、また過去分詞の由来動詞が語彙的に完了相(telic)であるのか未完了相(atelic)であるのかによって、それらの組み合わせに一定の制限のあることが(2)のように、すでに Gili Gaya (1973) などにより指摘されている。

(2) a. *La puerta había sido abierta.* 《完了時制+完了相動詞》

そのドアは開けられていた。

b. *(Fulano) fue querido (por todos los que le conocieron).*

《完了時制+未完了相動詞》

(誰かが彼を知る人たち皆に) 愛されていた。

c. **La puerta es abierta por el portero.* 《未完了時制+完了相動詞》

そのドアは守衛によって開けられる。

d. *Fulano es (era) muy conocido en aquella comarca.*

《未完了時制+未完了相動詞》

(誰かは) その地方で知られている (いた) .

すなわち、完了時制では完了相動詞および未完了相動詞と組むが、未完了時制では未完了相動詞とのみ親和性を見せる⁴のである。そこで、コーパスによる両受動文の生産性調査では両受動文がとる時制は(3)のように、完了時制と未完了時制とに分類することにした。

(3) a. 完了時制：点過去，現在完了，過去完了，未来完了，接続法現在完了，接続法過去完了。

b. 未完了時制：現在，線過去，未来，過去未来，接続法現在，接続法過去。

また過去分詞となるもとの動詞の語彙アスペクト(Aktionsart)は Vendler (1967)にしたがって(4)のように4つのグループに分けた。

(4) a. 到達動詞：descubrir 発見する，detectar 見つける，recibir 受け取る

b. 達成動詞：construir 建築する，destruir 破壊する，quemar 燃やす

c. 状態動詞：saber 知る，conocer 知る；escuchar 聞く，ver 見る

d. 活動動詞：buscar 探す，llevar(corbata) (ネクタイを) 身につける；estudiar (derecho)

(法学を) 勉強する，escribir (cartas) (手紙を) 書く，leer (cartas) (手紙を) 読む

到達動詞と達成動詞は完了相である。到達動詞(achievement verbs)は一つの限界点により特徴づけられるのに対し，達成動詞(accomplishment verbs)は限界点と継続部からなる。一方，広義に捉えた状態動詞と主体の意志性を含意する活動動詞(action verbs) はともに未完了相(atelic)動詞と分類される。

このような2つの時制アスペクトと4種の語彙アスペクトを基準にコーパス中のSER受動文とSE受動文の出現頻度・比率を調べた。例文末尾の〈〉内はコーパスのタイトル略号(注7を参照)である。

ここではわずか2例だけであるが，完了相から到達動詞のdetectar，また未完了相から活動動詞buscarの例を紹介しておこう。(5)(6)のようにそれぞれ4つの組み合わせの例文が可能になる(詳細については上述の研究を参照)。

(5) a. Nueve embarcaciones del mismo tipo fueron detectadas...por el servicio

Marítimo. 〈95〉同種の9隻の船が沿岸警備隊により発見された。

b. 例なし。

c. Pedro Solar no tiene ni idea de por qué no se detectó el fallo del vaporizador...

〈95〉 ペドロ・ソラールには気化器の欠陥がなぜ見つけられなかったか全くわからない。

d. ...en los hechos se detectaban las vacilaciones del Gobierno al respecto...

〈95〉 それらの事実の中にこの問題に関する政府のためらいが見てとられた。

(6) a. 例なし。

b. ...dos de ellos son buscados por asesinato en sus respectivos países, Ghana y Liberia. 〈95〉 彼らのうちの2人は殺人罪でガーナ、リベリアのそれぞれの国で捜索されている。

c. Para los cursos de Bachiller se buscó un profesorado joven y enérgico... 〈10〉 中等教育のために若くて精力的な教員が求人された。

d. Cuando un amor se acaba, se busca el sustituto. 〈co〉

一つの愛が終わると、その代わりになるものが求められる。

ともに、(a)文はSER受動文が完了時制で用いられた場合、(b)は未完了時制、(c)はSE受動文が完了時制で、また(d)文はそれが未完了時制で用いられた場合の組み合わせになり、(5)では(b)の例、(6)では(a)の例が検出できなかった。

1.3 調査結果

それぞれの動詞について、次表のような生起結果が得られた。

	SER 受動文		小計	SE 受動文		小計	合計
	完了時制	未完了時制		完了時制	未完了時制		
detectar 完了相	5 100%	0 0%	5 (20%)	16 80%	4 20%	20 (80%)	25
buscar 未完了相	0 0%	2 100%	2 (6.2%)	6 20%	24 80%	30 (93.8%)	32

detectar と buscar の受動文生起数

1) この表からわかることは、まず、両受動文の生起数の中に占める SE 受動文の比率の圧倒的な高さである。完了相動詞 *detectar* の SER 受動文は 5 例で 20%、SE 受動文は 20 例で 80% を占めている。また、未完了相動詞の *buscar* でも SER 受動文の 2 例がわずか 6.2% であるのに対し、SE 受動文は 30 例、93.8% にもなっている。完了相の達成動詞グループ (例, *construir*) は *detectar* と、また未完了相の状態動詞グループ (例, *saber*) は *buscar* と共通した傾向を見せることが調査からわかった。これは、1.1 節でふれた直感を裏づけることになるだろう。すなわち、スペイン語の 2 種の受動文のうち SE 受動文が SER 受動文に侵食している可能性を示唆している。

2) SER 受動文のAspect制約について見てみると、時制Aspectと語彙Aspectの 2 種類が受動文形成に一定の制約を課していることがわかる。上の表のように完了相の *detectar* は完了時制に、また未完了相の *buscar* は未完了時制に集中している。その他の動詞グループについても、SER 受動文の語彙Aspectごとの時制Aspectとの相関を調べた結果、完了相と完了時制、未完了相と未完了時制の適合性が、概ね、次表のようになっていることがわかった。

	完了時制	未完了時制	語彙Aspect
到達動詞 (<i>detectar</i>)	√		完了相
達成動詞 (<i>construir</i>)	√		完了相
状態動詞 (<i>saber</i>)		√	未完了相
活動動詞 (<i>buscar</i>)		√	未完了相

SER 受動文のAspect相関

3) さらに、SE 受動文についても、*detectar* と *buscar* の比較からわかるように、完了相の到達動詞で完了時制の比率が高い⁵ こと、反対に未完了相の状態動詞では未完了時制の比率がきわめて高いという相関が認められた。

この節では、コーパスに基づいて、スペイン語に見られる 2 種類の受動文の生産性調査の概略を紹介したが、以下ではここで用いられた現代スペイン語コーパスについて述べていこう。

2. KLM Corpus について

SER 受動文と SE 受動文を抽出するために使用されたコーパス (KLM Corpus) は, 約 35.5 メガバイト, 総語数 約 578 万語 の小規模のスペイン語コーパスであるが, 現代スペインの新聞, 小説, と会話戯曲 の3ジャンルのほぼ同量のテキストから成っている. 新聞はスペインの日刊紙 El Mundo (1995)⁶. 小説は, Eduardo Mendoza <La ciudad de los prodigios> や Carmen Martín Gaité <Nubosidad variable>など 21 タイトル. 会話と戯曲は, M. Esguevara y M. Cantarero <El habla de la Ciudad de Madrid. Materiales para su estudio> など, 録音会話からテープおこしされたテキストと, Ana Diosdado <Anillos de oro> や Antonio Gala <Carmen Carmen> などのシナリオ 35 タイトル. 計, 56 タイトルから成る⁷. 以下の例文で分かるように, レコードの冒頭にはコーパスのタイトル名を示す略号が付与されている. また, según の ú のようなスペイン語の特殊文字は, u' のように2バイトで表示されている:

...(前略)...

06: Que es aficionado a la caza, segu'n se deduce de los rifles y trofeos colgados en la pared. Entre ellos destaca la magnífica cabeza disecada de un <Nyata Mangoma>, una rari'sima variedad de bu'falo africano.

...(中略)...

06: CLAUDIO.- Un <nyata mangoma>, una especie casi extinguida de bu'falo-cafre. Cuando un macho se hace viejo, es expulsado de la manada. Pero e'l la sigue y cuando llega la e'poca del celo, lucha por la hembra con ejemplares ma's jo'venes. Sabe que sera' vencido, pero no renuncia.

...(後略)...

3. SER 受動文の検索と分析データの作成

対象コーパスから SER 受動文の候補を抽出する方法としては, 例えば, a. 検索スクリプトを利用して, 直接, SER 動詞の活用語形 に [ai]d[oa](s) に代表される過去分詞語形で終る語が続くレコードを取り出す, b. テキストに SER 動詞の活用語形と過去分詞の語形データを付与してから, 検索スクリプトで取

り出す，という方法が考えられる．a. では，「SER 動詞＋過去分詞」に特化した検索スクリプトを書くのでなければ，f50401⁸のような汎用の検索スクリプトに100余りのSER動詞の活用語形を代入することになる．また，過去分詞語形の候補としては，[ai](')d[oa](s)だけでなく，puesto, hecho, impresoなども想定して，[silro]t[oa](s), [ei]ch[oa](s), es[oa](s)などで終る語も検索語に含めることになり，SER受動文のヒット率がかなり低下する．従って，高垣・宮本(2003)では，b.の方法を採用した．

3.1 コーパスへのSER受動のデータ付与

b.の手順は以下のとおりである：

(7) 十分に大きな単語リストから約7500の動詞(不定詞)のリストを作る⁹：

abacorar(vt)

abajar(vt.vi)

...(中略)...

zurear(vi)

zurrar(vt)

(8) 不定詞形から，過去分詞も含め活用語形を展開した約75万2千の動詞活用語形リストを作る¹⁰：

abacorar(abacorar.0.inf.0)

abacorando(abacorar.0.ger.0)

abacorado(abacorar.0.pp.0)

abacoro(abacorar.i.p.1s)

...(中略)...

habri'an zurrado(zurrar.i.cp.3p)

haber zurrado(zurrar.0.infp.0)

habiendo zurrado(zurrar.0.gerp.0)

zurrad(zurrar.m.p.2p)

(9) (8)のリストから過去分詞を取り出し，過去分詞(男性単数形)リスト

を作る:

abacorado(abacorar.0.pp.0)

abajado(abajar.0.pp.0)

...(中略)...

zureado(zurear.0.pp.0)

zurrado(zurrar.0.pp.0)

(10) (8) のリストから SER 動詞の活用語形を取り出し, 103 個から成る SER 動詞の活用語形リストを作る:

ser(ser.0.inf.0)

siendo(ser.0.ger.0)

sido(ser.0.pp.0)

soy(ser.i.p.1s)

...(中略)...

habri'an sido(ser.i.cp.3p)

haber sido(ser.0.infp.0)

habiendo sido(ser.0.gerp.0)

sed(ser.m.p.2s)

(11) (9) のリストに, 過去分詞の女性形, 複数形を加えて, 過去分詞リストを作る:

abacorado(abacorar.0.pp.0)

abacorada(abacorar.0.pp.0)

abacorados(abacorar.0.pp.0)

abacoradas(abacorar.0.pp.0)

...(中略)...

zurrado(zurrar.0.pp.0)

zurrada(zurrar.0.pp.0)

zurrados(zurrar.0.pp.0)

zurradas(zurrar.0.pp.0)

3. 2 追加処理

さらに、SER受動文の分析に役立ちそうな、以下のような処理も試みた：

(14) a. SER受動文候補の ser の文法データ(法, 時制, 人称/数)の頻度数によるリスト¹³を作る：

```
1305 i ind 3s
1114 0 inf 0
511 i ind 3p
475 i p 3s
... (中略)...
1 i pl 1p
1 s p 2p
1 s pre 1s
1 s pse 2s
```

(14) b. SER受動文候補の過去分詞を不定詞に戻した頻度数リストを作る：

```
135 detener
85 condenar
75 asesinar
... (中略)...
1 volver
1 vomitar
1 zarandear
```

(14) c. 前後幅 1 0 語で、現在分詞も含めて、不定形のSER受動文候補を取り出す¹⁴：

```
06:: CAROLINA. - !Pues hijita, ya hay que [ser(ser.0.inf.0)] [despistada!(despistar.0.pp.0)]
06: ma' rmol fuese (ser.s.pse.1s) (ser.s.pse.3s) oro.
27:: Y au'n asi' no sera' (ser.i.f.3s) digno de [ser(ser.0.inf.0)] [pisado(pisar.0.pp.0)] por Su
Majestad.
```

...(中略)...

95:: 95: Este, al [ser(ser.0.inf.0)] [cuestionado(cuestionar.0.pp.0)]
sobre la posibilidad de poder contar con Mike Smith para
95:: los ocho an~os, no le ha
ocurrido (ocurrir. 0.pp. 0) nada dignode [ser(ser.0.inf.0)] [recordado.(recordar.0.pp.0)] Y
Carmencita explica que, para crecer, hay que olvidar. Pero,

(14) d. SER(不定詞)受動の直前の1語の頻度数リストを作る :

190 de
120 a
65 para
59 que
55 al
55 puede
...(中略)..
1 temiendo
1 tolera
1 valido
1 yo

(14) e. SER(不定詞)受動の前にくる動詞の文法データ(法, 時制, 人称/数)による頻度数のリストを作る :

131 i p 3s
53 i p 3p
36 i imp 1s
...(中略)..
1 i f 1s
1 i p 2s
1 s pse 3p

sigue y cuando llega

...(中略)...

95: ya se' que
esto no es divertido. Pero ¿que' quiere? [Soy] [jubilado] y esto tiene algo de
especta'culo. Al fin y al

95: fa'cil. Quedan socialistas
de base, histo' ricos sin sobre, pero hemos [sido] [estafados] por una gigantesca
improvisacio'n, desde los civilones conversos a los

(15) b. (15a) のファイルを目視チェック用に、過去分詞でソートする：

[abandonada.] se: menos.

El asunto se me iba de las manos. Acte' [fue] [abandonada.] Nero'n empenzo' a salir solo
de noche, como un jovencito

[abandonada] 95: 95: Afirma que [fue] [abandonada] por sus familiares y ha
quedado ingresada en un centro

...(中略)...

[vuelto] t: eran
de diez o doce mil pesetas al cambio, o [sea,] [vuelto] a pesetas.

[zarandeado] 95: la escala de Richter.

Desde el martes, este archipie' lago ha [sido] [zarandeado] por cientos de re'plicas.

(15) b'. このファイルを目視チェックして、例えば、[vuelto] のレコードは、o [sea,]
[vuelto] なので、SER受動文ではないから、削除する。このようにして「SER受動文
ファイル」を作る。

(15) c. この(15b). の「SER受動文ファイル」を、時制別で大きくソートし、さらに、
それぞれの中は過去分詞でソートしたファイルを、再チェックの意味も込めて、作
成する：

{0.gerp}[elegido] 95: ocupaba en el Ejecutivo de Berlusconi, ha declarado que

<habiendo [sido] [elegido] ministro del Tesoro por el Polo de la Libertad, me

...(中略)...

{i.ind}[abandonada.] se: menos. El asunto se me iba de las manos. Acte' [fue]
[abandonada.] Nero'n empezo' a salir solo de noche, como un jovencito

{i.ind}[abandonada] 95: Afirma que [fue] [abandonada] por sus familiares y ha
quedado ingresada en un centro

...(中略)...

{i.ind}[vistos] 95: audiencia de 9,4 millones, en la u'ltima noche de 1991 [fueron]
[vistos] por 9,7 millones y 1992, por 8 millones. En 1993

{i.ind}[votado] 1: Un subsidio [fue] [votado] al punto.

...(中略)....

{s.p}[votada] 95: llamado la Iniciativa de Derechos Civiles de California, para que
[sea] [votada] en todo el estado en 1996, fecha en la que

この (15b') と (15c) のファイルを利用して、1 節で述べたような SER 受動文の研究を行った。

4. SE 受動文の検索と分析データの作成

SER 受動文の場合と同様に、SE 受動文の候補を取り出す。

4.1 SE 受動文の動詞分類

高垣・宮本 (2003) では、SE 受動文の調査用動詞を 6 つのグループに分類した¹⁶。

(16) a. 人間動作主を前提とし、受動解釈が自然な 18 動詞：

construir, cerrar, recibir, reconocer, detectar, aprobar, suspender, rechazar,
instalar, colocar, comprar, vender (完了相) ;

escribir (cartas), leer (libros), pintar, estudiar, llevar, conducir (本来未完
了相であるが、escribir una carta, leer el libro のように限界性の直接目
的語を伴って完了相に転じると考えられるもの)

b. 自動詞化する可能性がある 16 動詞 :

destruir, encontrar, descubrir, abrir, detener, quemar, hacer, poner,

abandonar, seguir (完了相)

ver, escuchar, oír, entender, saber, conocer (未完了相の状態動詞)

c. 人間主語で受動文になる可能性のある 3 動詞 :

buscar, aplaudir, necesitar

d. 人間主語で受動文になりにくいと思われる 2 動詞 :

asesinar, atacar

e. 「por + 動作主」をとりうる 1 動詞 :

publicar

f. 二重目的語をとる 1 動詞 :

enviar

SE受動文の候補は、次ぎの4つのパターンを抽出した。非人称、(能動)再帰、相互用法など、SE受動文以外と区別できるように、前後幅は、広く、25語とする。以下、動詞 escribir を例に説明する :

(17) a. se_escribir: se が動詞の前に来るパターン候補を取り出す。se と動詞の間は最大2語¹⁷で抽出し、KWIC整形する :

ci: esmeraldas
y te haces con dineros para el viaje. A Domingo dale mi tabardo que tanto le gustaba.
El libro, para todos, porque para todos [se] [escribio'...](Lleva'ndose las manos de
OLALLA a su cara.) Tus manos. Siento que todo me llego' y me llega a trave's de tus manos.
(OLALLA con

...(中略)...

95: que <no data>, como de si' mismo
deci'a a veces don Pi'o Baroja, con aquella ironi'a de aldeano internacional,
<archieuropeo>, como e' l mismo se definio'. [Se] [escribe] mucho de vi'ctimas y
verdugos, estamos llegando al estragamiento por exceso, pero nadie contribuye a devolverles
a los vascos lo suyo (el Guernica, por ejemplo)

(17) b. se_v_escribir: se と動詞(escribir/escribiendo)の間にもう一つ動詞が来るパターン候補を取り出す。se と動詞(escribir/escribiendo)の間は最低1語, 最大4語¹⁸で抽出し, KWIC整形する:

44: sitios prohibidos.
En cambio la Loli, por ejemplo los tiene todos en sitios visibles o <tolerados para menores>. Los mi'os <toos> en el nu'mero cuatro [(Se) ve a CARLOS [escribiendo] en su mesa de despacho) OCTAVIO.- (Leyendo lo escrito) Estimado Carlos: Como no creo que quieras a Carmen ni piensas casarte, por la presente te
...(中略)...

95: El final,
escrito a sangre y fuego en la memoria de la Historia, no puede ser sino e'pico. Los personajes lloran,sucumben. Hasta Steven Runciman [se] emociona al [escribir] sobre esas u'ltimas horas de la capital de Bizancio. Otros ingleses de su escuela historiogra'fica, al abordar temas que les fueron muy caros, si' consiguieron

(17) c. escribir_se: se が動詞(escribir/escribiendo)に後接する¹⁹パターン候補を取り出す:

ue: ue :
ANDRE'S.—Ya ves, eso no lo comprendo:[escribirse] con personas desconocidas.
10: 10: -- Supongo que tiene derecho a [escribirse] con quien quiera, ?no?
...(後略)...

(17) d. escribir_se_p: se が完了形の動詞(haber escrito/habiendo escrito)と組むパターン候補を取り出す:

例なし²⁰

(18). チェック用も兼ねて, それぞれの頻度リストを作る:

(18) a. se_escribir 候補の頻度リスト:

21 se escribe
 5 se escriben
 5 se escribi'a
 ...(中略)...
 1 se hizo habi'a escrito
 1 se la habi'a escrito
 1 se santiguo' y escribio'

(18) b. se_v_escribir 候補の頻度リスト :

2 se puede escribir
 2 se puso a escribir
 1 se contenta con escribir
 ...(中略)...
 1 se ve a carlos²¹ escribiendo
 1 se vei'a la ma'quina de escribir
 1 se vio forzado a escribir

(18) c. escribir_se 候補の頻度リスト :

1 derecho a escribirse
 1 intimidad sexual escribie'ndose
 1 lo comprendo escribirse
 1 podri'a seguir escribie'ndose
 1 que podri'a escribirse

(18) d. escribir_se_p 候補の頻度リスト :

該当なし²²

(17a) から (18d) までの処理を, (16a) から (16f) に挙げた 41 個の動詞について, シェルスクリプトで一括して行った²³. そして, (17a) から (17d) のパターン別に抽出したファイルを利用して, 1 節で述べたような S E R

受動文と S E 受動文の対比研究を行った。

むすび

コーパスを対象に、動詞の活用展開、データ付与、検索などのスクリプトを駆使することで、従来は困難であった大量のテキストを対象とする文法テーマの研究が可能になる。本稿では、スペイン語の受動文の研究のための、コーパス処理の手順について述べたが、ここで扱った受動文をめぐることは、コーパスを利用して今後さらに次のような研究課題の可能性が残されている。

まず、SER 受動文については、主語の特性による生起を調べる必要があるだろう。SE 受動文が基本的には事物しか許容しないのに対し、SER 受動は人間主語でもよいとされる。しかし、実際の生起率はどの程度なのだろうか。とくに 1 人称、2 人称主語の SER 受動文はそれほど頻度が高くないように思える。さらに、動作主を表す *por* (ないしは *de*) 句を現実にはどの割合で伴っているのか検証してみることも可能であろう。

一方、S E 受動文では、主語位置が動詞後であることが無標とされる。このような場合主語と対比的に場所補語が動詞に先行する傾向が予想される。このような語順の本格的調査も今後の課題となるだろう²⁴。さらには、S E 受動文についても、原則的には伴われないことになっている動作主の *por* 句の生起についても確認しておく必要がある。

さいごに、両受動文に対応する能動文との生起数比較がある。試みに *buscar* を含む能動文を同じコーパスで検索してみると、能動文は 1 2 1 5 例見つかった。上表より両受動文が計 3 2 例なのでわずかに 2.3% を占めるのみで、受動文の生産性はきわめて低いことになる。その他の動詞についてもこのような比率を確かめることにより、冒頭で問題にしたスペイン語受動文の低い生産性の実態が明らかになるであろう。

注

- 1) 諸言語での研究については, Kawaguchi et al.(2003) を, スペイン語のデジタル・コーパスについては, Moreno Fernández(2003)を, コーパスを用いたスペイン語文法の最近の研究例としては, 木越(2004), 結城(2004), Miyamoto(2003), 大家(2003), 宮本(2002)などを参照.
- 2) AWK, Perl, シェルスクリプトについては, それぞれ, 例えば, エイホ他(1989), ウオール他((1997), プリン(2003)などを参照.
- 3) Fenández Ramírez(1986), De Miguel(1992)などを参照. しかし, これら両タイプの受動文が現実にはいかなる生産性をもっており, どのような比率で使い分けられているのかなどについて明示的な記述がなされているとはいえないのが現状である. この点, 英語の受動文については大量コーパスを用いて文体レベルによる使用率まで示してある Biber (1999)のような啓蒙的文法書がめずらしくない. スペイン語文法についても同種の調査・分析が急務となる. 高垣・宮本(2003)や高垣(2004), Takagaki(to appear)などのコーパスによる受動文研究はこのような試みの一つ言えるだろう.
- 4) (2c)は習慣の行為を表すのであれば許容される.
- 5) 但し, 完了相の達成動詞については, 未完了時制の比率の方が幾分高いという結果が出ている.
- 6) El Mundo 1995, Primer Semestre, CD-ROM, Unidad Editorial, S.A. から約 13.4メガの量を抽出した.
- 7) 会話戯曲には, 上田博人氏の <Análisis lingüístico de las obras teatrales españolas> のシナリオテキストも含む. KLM Corpus のタイトル名については, 高垣(2003), Takagaki (to appear)を参照.
- 8) f50401 の仕様は, 山田(2003:15-16)を参照. 山田は f50401 に語間の最小値の機能も加えたものを Perl に移植した.
- 9) () 内は動詞のデータ: vt 他動詞, vi 自動詞.
- 10) この動詞活用形を展開する宮本のスクリプトのアルゴリズムについては, 大家(2003:31-32)の解説が詳しい. そこで説明されている y00316 は, 今回作成したシェルスクリプト y11103.sh の動詞活用形を展開する部分に相当

する。()内のデータ: inf 不定詞, ger 現在分詞, pp 過去分詞, i 直説法, p 現在, 1s 1人称単数, cp 過去未来, 3p 3人称複数, infp 不定詞完了形, gerp 現在分詞完了形, m 命令法.

1 0 bis) 動詞活用形のタグ付けスクリプト y00627 を使用した. 形容詞, 名詞のタグ付けスクリプトについては, 宮本(1997:97-98) を参照.

1 1) () 内のデータ: f 未来.

1 2) もちろん, 前後に 10 語以上ある場合で, それ以下の場合, ある語数だけ取り出す. 左端の数字は, タイトルの略号. [...] で囲まれているのが検索対象のキーワード.

1 3) 以下の記号: ind 点過去, pl 過去完了, pre (接続法)未来, pse (接続法)過去 SE 形.

1 4) 不定形 SER の受動文については, 高垣(2003)に詳細な分析がある.

1 5) シェルスクリプト y11103.sh は, DynaBook G7(プロセッサ: モバイル Pentium 4 1.90 GHz-M, メモリ: 512 MB) の TurboLinux 8 上で, この処理に約4分20秒要した.

1 6) ただし, このグルーピングは調査とともにすべてが妥当なものでないと判明したものも含まれるので注意されたい. 例えば, abandonar.seguir が自動詞化する動詞グループ(16b)に含めるとすべきではなかった.

1 7) se le ha escrito のようなパターンを想定して, se と動詞の語間は, 最大2語とする.

1 8) (17a) のパターンの多くを排除するために, 語間に最低1語が入り, se le han empezado a escribir などと想定して, 最大4語とする. 語間に最小値と最大値を代入するので, 検索スクリプトは, f50401 ではなく, 語間の最小値も指定する機能を持たせた改良版の y30104 を使用する. y30104 については, Miyamoto(2003:105-108) を参照.

1 9) escribase のような, 肯定命令形に後接する場合も想定している.

2 0) このパターンの実例(動詞 vender)を挙げる:

1: 1:A veces este proceso no se deteni'a al [haberse] [vendido] ya todas las viviendas de la zona, sino que empezaba entonces una segunda ronda de reventas y hasta una tercera y una cuarta.

95: 95: Otros, porque le acusan de haber traicionado el espíritu liberal y [haberse] [vendido] descaradamente. A fin de cuentas, la última palabra la tienen los americanos. Y si hacemos caso de la encuesta telefónica realizada por la CNN nada

2 1) 頻度リストでは、文字はすべて、小文字化されている。

2 2) このパターンの実例(動詞 vender) : 2 haberse vendido

2 3) 注 14) で述べたシステムで、約 2分 39秒要した。

2 4) 結城 (2004) を参照。

対象コーパス

KLM Corpus(2001).

参考文献

Biber, Douglas et al. (1999) *Longman Grammar of Spoken and Written English*.
Longman.

De Miguel, Elena (1992) *El aspecto en la sintaxis: Perfectividad e impersonalidad*.
Ediciones de la Universidad Autónoma de Madrid.

Fernández Ramírez, Salvador. (1986) *Gramática Española. 4. El verbo y la Oración*.
Arco Libros.

Gili Gaya, Samuel (1973) *Curso Superior de Sintaxis Española* (onceava edición).
Vox.

Kawaguchi, Yuji, Susumu Zaima, Toshihiro Takagaki, Kohji Shibano & Mayumi
Usami (2003) *Proceedings of the First International Conference on Linguistic
Informatics*. Tokyo University of Foreign Studies.

Miyamoto, Masami (2003) "A Formal Analysis of Spanish Adjective Position",
Kawaguchi et al. (2003), pp.93-108.

Moreno Fernández, Francisco (2003) "Corpus of Spoken Spanish Language. The
Representativeness Issue", Kawaguchi et al. (2003), pp.47-63.

Takagaki, Toshihiro (to appear) "On the productivity of the Spanish passive

- constructions” *Linguistic Informatics II, Linguistic Studies 1, Corpus-Based Analyses on Sentence Structures*. Tokyo University of Foreign Studies.
- Vendler, Zeno (1967) *Linguistics in Philosophy*. Cornell University Press.
- ウォール, ラリー, トム クリスチャンセン, ランダル L. シュワルツ(1997)『プログラミング Perl 改訂版』, オライリージャパン.
- エイホ, A.V., B.W. カーニハン, P.J. ワインバーガー(1989)『プログラミング言語 AWK』, トッパン.
- 大家正登(2003)「西語受動文の適格性」(卒業論文), 東京外国語大学.
- 木越勉(2004)「スペイン語名詞句内の形容詞の位置」(修士論文), 東京外国語大学.
- 布林, ブルース(2003)『入門 UNIX シェルプログラミング(改訂第2版)』, ソフトバンク.
- 高垣敏博(2003)「不定形 ser による受動文」『スペイン語学研究』18, pp.101-121.
- 高垣敏博・宮本正美(2003)「スペイン語受動文の生産性について—コーパスをもとに—」, (SELE 2003, 口頭発表, 8月27日, 伊豆長岡).
- 高垣敏博(2004)「スペイン語受動文の生産性について」『スペイン語学論集』, くろしお出版.
- 宮本正美(1997)「ABCにおける形容詞の位置」, 神戸外大論叢, 48:3, pp.77-98.
- (2002)「スペイン語における新語抽出—2—: (語幹的) 接尾辞」, 『神戸外大論叢』53:6, pp.27-42.
- 山田純子(2003)「名詞句の修飾構造 ポルトガル語形容詞の修飾方向とその性質の研究」(卒業論文), 東京外国語大学.
- 結城健太郎(2004)「スペイン語の語順—特に移動動詞が含まれる文について」(修士論文) 東京外国語大学.

Retrieval of Passive Constructions in Spanish Corpus

Masami MIYAMOTO & Toshihiro TAKAGAKI

This paper presents a semi-automatic extraction procedure for passive constructions in a Spanish corpus and a brief analysis of the two types of Spanish passive constructions formed with <SER + past participle> and <reflexive pronoun SE >, with special reference to their productivity.

In order to analyze Spanish passive constructions using texts called the <KLM Corpus>, we adopted the technique of tagging the conjugated forms of Spanish verbs (including past participles) to the corpus and extracting the candidate texts of <SER> and <SE> passive sentences from the tagged corpus. We have made this processing semi-automatic by various (shell) scripts of our own, such as a script for deriving conjugated forms of Spanish verbs, a script for tagging these conjugated forms to a corpus, a retrieval script, etc. which are all written in AWK or Perl and run on Linux.

Our analysis of the <SER> and <SE> passive sentence candidates extracted by the above procedure has come to the following conclusions:

- 1) <SE> passive sentences are found to be more productive than <SER> passive sentences, as expected.
- 2) With regard to <SER> passive sentences, a clear correlation between the lexical aspect and the inflexional aspect of verbs can be observed, that is, <telic> verbs are preferred in perfective tenses whereas <atelic> verbs are dominant in imperfective tenses.
- 3) <SE> passive sentences also show a similar correlation between lexical and inflexional aspects, as shown characteristically in the cases of achievement (telic) verbs and state (atelic) verbs.