

Implantación en máquina de la interlengua : nombres cuyos rasgos flexivos de género y número no coinciden con el español estándar.

Bárbara Méndez

Universidad Nacional de Rosario
Facultad de Humanidades y Artes
Rosario, Argentina
barbaramendez555@hotmail.com

Abstract:

In this work I propose to realize the automatic recognition of the interlanguage of learners of Spanish as L2, specifically we treat the names which inflectional features of gender and number are different from the standard language. It is necessary to add that I develop issues that I worked in the article " Automatic analysis of the interlanguage: gender and number assignment different from the standard language in the nucleus noun sintagma (snn). That is, departing from the mentioned work, my goal is to deepen the study of the names that presents these features, so we will study the nucleus noun sintagma (snn) and we consider some cases of noun sintagma (SN). Then, we make the implantation in machine that allows the recognition of the snn and the SN.

Regarding the implantation in machine, we use two tools, Smorph and Mps. We use one of the methods proposed in the mentioned study, and we realize the necessary changes in order that the new cases of snn and Sn can be extracted it.

Keywords: Interlanguage, implantation in machine, names which inflectional features of gender and number are different from the standard language.

Resumen:

En este trabajo me propongo realizar el reconocimiento automático de la interlengua de aprendientes de español como L2, específicamente se tratarán los nombres que presentan rasgos de género y número diferentes a la lengua estándar.

Cabe agregar que retomo y desarrollo cuestiones que trabajé en el artículo "Análisis automático de la interlengua: asignación de género y número diferentes a la lengua estándar en el sintagma nominal núcleo (snn). Es decir que, partiendo del trabajo mencionado, mi objetivo es ahondar en el estudio de los nombres que presentan estas particularidades, estudiar los sintagmas nominales núcleos (snn) y considerar algunos casos de sintagmas nominales (SN). Luego, realizar la implantación en máquina que permita el reconocimiento tanto de los snn como de los SN.

En lo que respecta al reconocimiento automático, se hace uso de dos herramientas informáticas, Smorph y Mps. Utilizo uno de los métodos propuestos en el mencionado estudio, y realizo los cambios necesarios para que se puedan extraer estos nuevos casos de snni y de SN.

Palabras claves: interlengua, implantación en máquina, nombres con rasgos de género y número diferentes de la lengua estándar.

1. INTRODUCCIÓN

En este trabajo me propongo realizar el reconocimiento automático de la interlengua de aprendientes de español como L2, específicamente se tratarán los nombres que presentan rasgos de género y número diferentes a la lengua estándar.

Cabe agregar que retomo y desarrollo cuestiones que trabajé en el artículo “Análisis automático de la interlengua: asignación de género y número diferentes a la lengua estándar en el sintagma nominal núcleo (snn) [1]. Es decir que, partiendo de cuestiones estudiadas en el trabajo mencionado, mi objetivo es ahondar en el estudio de los nombres que presentan estas particularidades, estudiar los sintagmas nominales núcleos (snn) y considerar algunos casos de sintagmas nominales (SN). Luego, realizar la implantación en máquina que permita el reconocimiento tanto de los snn como de los SN.

En primer lugar, se trabajarán nuevos casos de nombres de interlengua (ni), lo que permite ampliar la clasificación de éstos. En segundo lugar, se estudiarán otras estructuras de snni, en el trabajo mencionado se trabajó con los snni conformados por det + n ; se trabajarán aquí las siguientes estructuras de snni: det + adj + n y det + adv + adj + n. También se considerará el sintagma nominal, específicamente, los formados por: det + n + adj.¹

Por último, en lo que respecta a la implantación en máquina, se hace uso de dos herramientas informáticas, Smorph y Mps; se utiliza uno de los métodos propuestos en el mencionado estudio, y se realizan los cambios necesarios para que se puedan extraer estos nuevos casos de snni y de sn.

Se dispone de un corpus que consta de textos escritos de aprendientes de español como L2 que han terminado el nivel inicial y son hablantes de distintas L1 (francés, holandés, alemán, inglés).

2. LA FLEXIÓN NOMINAL

En términos generales, el nombre en español puede presentar cuatro formas: cuando posee los rasgos flexivos de género y número, como *gato*, *gatos*, *gata*, *gatas*; y dos formas cuando sólo admite la oposición de número, como *libro*, *libros*.

Siguiendo a Ambadiang [2] se pueden señalar marcas canónicamente asociadas con los rasgos flexión de género y número. En lo respecta al género, la distinción entre masculino y femenino se reconoce principalmente por la oposición *-o / -a* (*perro / perra*), *-e/o* (*jefe/jefa*) o la ausencia del rasgo flexivo para el masculino (*león / leona*).

En otros casos, la discriminación entre los dos géneros sólo es posible gracias al artículo o el resto de las concordancias, como en *el / la artista*.

En lo que se refiere al número, la diferencia entre singular y plural suele corresponderse con la oposición $\emptyset / -s$, $\emptyset -es$: *perro / perros*, *león / leones*. Cabe agregar que en algunos nombres, la distinción se manifiesta sólo por el artículo, como en *el / los atlas*. Centrándonos en la problemática planteada, la asignación de género y número a nombres diferentes a la lengua estándar, podemos señalar ciertos casos donde reconocemos la posible tendencia de los aprendientes de español como L2 a asignar un género o un número que no es que corresponde.

Por un lado, se trabaja el caso de los nombres cuyas marcas flexivas de género no se asocian con las marcas canónicas:

¹ Aclaraciones: det (determinante), n (nombre), adj (adjetivo), adv (adverbio)

- (i) nombres femeninos terminados en –o, como *foto, moto*
- (ii) nombres masculinos terminados en –a, como *día, mapa, clima, cometa, el planeta*
- (iii) nombres que terminan en –e, como *elefante*
- (iv) nombres terminados en consonantes, con ausencia de marca de género, como *rey, posibilidad*.

Por otro lado, en lo que respecta a las marcas flexivas de número, cabe mencionar a:

- (v) los nombres con igual marca de número tanto para el singular como para el plural, como *crisis, lunes*.

3. NOMBRES DE INTERLENGUA (NI): CASOS MÁS FRECUENTES

Llamaremos nombres de interlengua (ni) a los nombres que presentan desviaciones con respecto a la lengua estándar. Como se dijo anteriormente, los ejemplos que analizaremos fueron extraídos del corpus de interlengua. A continuación, presento un cuadro con los casos de ni encontrados.

Tabla nº1: Casos de ni

Casos de ni	Ejemplos extraídos del corpus
nombres femeninos terminados en –o	(el) <i>radio, foto, moto</i>
nombres masculinos terminados en -a	(la) <i>sistema, problema, día, programa</i>
nombres masculinos terminados en –e a los que se les asigna el género femenino	(la) <i>viaje, paisaje,</i>
nombres femeninos terminados en –e a los que se les asigna el género masculino.	(el) <i>sangre, suerte, torre,</i>
Nombres terminados en consonante	
- femeninos terminados en –ción a los que se les asigna el género masculino	(el) <i>información, habitación,</i> (los) <i>vacaciones</i>
- femeninos terminados en –dad a los que se les asigna el género masculino.	(el) <i>posibilidad, universidad, ciudad</i>
nombres con igual marca de número tanto para el singular como el plural.	* <i>crisi, *lune</i>

3.1. El sintagma nominal (SN)

Es el grupo de palabras que forma un sintagma, cuyo núcleo está constituido por un nombre, un adjetivo sustantivado, un pronombre. En cuanto a la estructura interna del SN, el núcleo (nombre) determina la índole categorial del sintagma en su conjunto, así como también sus rasgos flexionales (género y número), como las características

sintácticas de la construcción, su complejidad, la compatibilidad con algunos tipos de especificadores y la posibilidad de selección de complementos.

Los siguientes ejemplos son extraídos del estudio sobre el SN realizado por Rigau [3]

- i. *Donó [su valiosa colección a este museo comarcal]*
- ii. *Quiere comprarse [el coche más caro]*
- iii. *Van a publicar [los buenos trabajos de mis alumnos]*
- iv. *[El deseo de llegar al polo norte]*
- v. *[El ataque aéreo]*
- vi. *[Los recién aprobados presupuestos]*

En líneas generales, los elementos que conforman el SN son:

- Determinantes (Det.) : artículos, demostrativos, indefinidos, numerales, interrogativos.

Det + N : *[el coche más caro]*

-Adjetivos o sintagmas adjetivales (SA)

Det + SA + N :

Det + N + SA: *[El ataque aéreo]*

-Preposiciones o Sintagmas Preposicionales (SP)

Det + N + SP: *[El deseo de llegar al polo norte]*

- Adverbios o Sintagmas adverbiales (SADV)

Det + SADV +SA + N : *[Los recién aprobados presupuestos]*

3.2. El sintagma nominal núcleo (snn)

Como puede verse, el estudio del SN implica el estudio de las relaciones internas de los elementos que conforman el sn, es decir tanto de aquellos elementos que van delante del núcleo como de los que se posponen a éste.

No obstante, a partir de la noción de snn nos limitaremos a considerar a las categorías que se encuentran delante del núcleo, es decir a izquierda de éste. Siguiendo a Solana y Rodrigo [3] podemos definir el snn como el segmento del sintagma que comienza con el inicio de éste y termina en su núcleo, como *la vieja casa* que se inicia con el determinante, y termina en su núcleo.

Retomando los ejemplos de SN, señalamos los snn en cada caso (en negrita).

- i. *Donó [**su valiosa colección** a este museo comarcal]*
- ii. *Quiere comprarse [**el coche** más caro]*
- iii. *Van a publicar [**los buenos trabajos** de mis alumnos]*
- iv. *[**El deseo** de llegar al polo norte]*

v. [*El ataque aéreo*]

vi. [*Los recién aprobados presupuestos*]

A partir de las categorías morfosintácticas que integran el snn, es posible determinar dónde comienzan, dónde terminan, cómo están compuestos y cuál es su núcleo. Las categorías se presentan en un orden, por lo tanto las combinaciones entre las categorías son limitadas. Estas regularidades y restricciones en cuanto a la combinación de las categorías determinan propiedades de linealidad y permiten su fácil extracción.

3.3. Casos encontrados en el corpus

Como se dijo anteriormente, me ocupo de los SN conformados por det + n + SA y, de los snn con las siguientes estructuras: det + n; det + adj + n.

A continuación, presento los casos de SN y snn presentes en el corpus que no coinciden con el español estándar, que llamaremos SNi, y snni.

- SNi

det + ni + SA

* *la día muy hermosa*

* *los gentes muy simpáticos*

- Snn

det + ni:

* *la sistema, *la problema, *el radio, *el foto, * el moto *los habitaciones *los informaciones, *los gentes, *el torre, *las paisajes, *el ciudad*

det + adj + ni:

* *las próximas días, *la seria problema, * el viejo radio*

det + adv + adj + ni

el muy lindo ciudad

4. IMPLANTACIÓN EN MÁQUINA

Para el análisis automático nos valdremos de Smorph y Mps, estas herramainetas trabajan en forma complementaria.

4.1 Smorph

Smorph es un segmentador y lematizador, un software especificado e implementado en el GRIL por Salah Ait-Mokhtar , bajo la dirección de Gabriel Bès. Es un analizador y generador morfológico que realiza la segmentación (tokenización) y el análisis morfológico en una sola etapa. Segmentar consiste en delimitar las palabras, en tanto que el análisis morfológico implica encontrar el lema y asignara cada unidad la información morfológica que le corresponde.

Smorph es una herramienta declarativa, por lo tanto debe declararse la información lingüística. La información utilizada está separada de la maquinaria algorítmica, esto hace que se la pueda adaptar al uso que quiera darse, con el mismo software se puede tratar cualquier lengua si se le cambia la información lingüística.

En Smorph debe declararse cinco tipos de informaciones:

Entradas: representan el diccionario lingüístico. Las palabras pueden aparecer con la indicación del modelo morfológicos, por ejemplo sistema @a1 o con la indicación de los rasgos, por ejemplo *muy/adv*.

Modelos: se señala la estructura morfológica. Como puede verse en el ejemplo anterior, @ introduce el modelo e indica el lugar en que va la raíz a la que se concatenan las terminaciones.

Por ejemplo para nombres del tipo *sistema* corresponde el modelo 1 que presenta la siguiente estructura:

@n1	-0	
+@		nom/masc/sg
+s		nom/masc/pl .

Terminaciones: diferentes terminaciones que cada lema puede presentar en su derivación morfológica.

Rasgos: para construir los modelos se recurre a rasgos morfológicos sintácticos (categoría, género, número, etc) Así por ejemplo se tienen: etiqueta morfosintáctica (EMS), género (gen) que tiene los valores femenino (fem) y masculino (masc).

Códigos ASCII: son una forma de expresar digitalmente el texto, se deben declarar los códigos que se desean utilizar, por ejemplo el espacio, el signo de interrogación, etc

4.2 MPS

Recibe en entrada una salida Smorph (en formato Prolog) y va a dar en salida otro formato según el analizador que se vaya a utilizar; MPS podrá modificar las estructuras de datos recibidos en la entrada. MPS ejecuta dos funciones principales: la Reconstrucción y la Correspondencia; la Reconstrucción a su vez puede ser de dos tipos diferentes, el Reagrupamiento y la División.

MPS es una herramienta declarativa, en donde mediante reglas se pueden expresar los valores de entrada (sobre dos o más estructuras de datos de la salida Smorph) y los valores de salida sobre la estructura reagrupada.

Las reglas que se declaren con la función de división van a provocar el efecto inverso. Son útiles para tratar las contracciones (por ej. una ocurrencia de del en español), afín de obtener en la salida una secuencia de entidades que sea análoga a las que Smorph asigna a las ocurrencias no contraídas en una cadena

Las reglas que se declaren con la función de correspondencia van a operar sobre una sola estructura de datos a la salida de Smorph y van a poder modificarla en otra estructura de datos. Estas reglas permiten formular en Smorph descripciones básicas, generales, y adaptarlas después a la exigencia de cada analizador o de cada aplicación, o enriquecerlas con nuevos pares de <etiqueta=valor>.

Cabe agregar que, como se verá en el siguiente apartado, hemos utilizado las reglas de reagrupamiento.

5. EXTRACCIÓN DE LOS SN Y DE LOS SNN:

Como se explicó en el punto anterior, Smorph realiza la segmentación y el análisis morfológico. A modo de ejemplo, presentamos la salida correspondiente al *ni sistema*: ['sistema', 'EMS', 'nom', 'GEN', 'masc', 'NUM', 'sg'].

Cabe agregar que trabajaremos con la información declarada para el español estándar.

Es decir, no realizaremos modificaciones en los ficheros mencionados.

En cuanto a Mps, en primer lugar, se crean reglas de reagrupamiento con los rasgos de concordancia. Esto permitirá extraer los SN y los snn que coinciden con el sistema lingüístico del español.

Regla nº1: art masc, sing + nom masc, sing + adj masc, sing → SN1

Ej.: *el edificio viejo*

Regla nº2: art masc, pl + nom masc, pl + adj masc, sing → SN2

Ej.: *los zapatos nuevos*

Regla nº3: art fem, sing + nom fem, sing + adj fem, sing → SN3

Ej.: *la calle oscura*

Regla nº4: art fem, pl + nom fem, pl + adj fem, pl → SN4

Ej.: *las escaleras anchas*

Regla nº1: art fem, sing + nom fem, sing → snn1

Ej.: *la casa*

Regla nº2: art masc sing + nom masc sing → snn

Ej.: *el afiche*

Regla nº3: art masc pl + ni masc pl → snn3

Ej.: *los autos*

Regla nº4: art fem sing + adj fem sing + ni fem sing → snn4

Ej.: *La vieja casa*

Regla nº5: art fem pl + adj fem pl + ni fem pl → snn

Ej.: *Las viejas casas*

Regla nº6: art masc sg + adj masc sg + ni masc sg → snn6

Ej.: *el interesante libro*

Regla nº7: art masc pl + adj masc pl + ni masc sg → snn7

Ej.: *los altos árboles*

En cuanto a los SNi y snni, el otro tipo de reglas que propongo son las siguientes:

Regla nº1: art fem sing + nom masc sing + adv + adj fem sing → snni1

Ej.: *La día muy hermosa*

Regla nº2: art masc pl + nom fem pl + adv + adj fem sing → snni1

Ej.: *La día muy hermosa*

Regla nº1: art fem sing + nom masc sing → snni1

Ej.: **la sistema, *la problema, *la viaje*

Regla nº2: art masc sing + nom fem sing → snni3

Ej.: **el sangre, *el posibilidad, *el universidad, *el radio*

Regla nº3: art masc pl + nom fem pl → snni2

Ej.: **los habitaciones, *los informaciones*

Regla nº4: art fem sing + adj fem sing + nom masc sing → snni4

Ej: **la seria problema*

Regla n°5: art masc sing + adj masc sing +nom fem sing → snni5

Ej: ** el viejo radio*

Regla n°6: art fem pl + adj fem pl +nom masc sing → snni6

Ej : **las próximas días*

Regla n°7: art masc pl + adj masc pl +nom fem sing → snni7

Ej: **los cómodos habitaciones*

Resultados de Mps:

- ['El edificio viejo', 'EMS', 'SN1'].
- ['Los zapatos nuevos', 'EMS', 'SN2'].
- ['La calle oscura', 'EMS', 'SN3'].
- ['Las escaleras anchas ', 'EMS', 'SN4'].
- [' la día muy hermosa ', 'EMS', 'SNi1'].
- [' los gentes muy simpáticos', 'EMS', 'SNi2'].
- ['la sistema', 'EMS', 'snni1'].
- ['el radio', 'EMS', 'snni2'].
- ['los habitaciones', 'EMS', 'snni3'].
- ['la serio problema', 'EMS', 'snni4'].
- ['el viejo radio', 'EMS', 'snni5'].
- ['las próximo día', 'EMS', 'snni6'].
- ['los cómodos habitaciones', 'EMS', 'snni7'].

5. CONCLUSIÓN:

En este trabajo hemos realizado la descripción de los nombres de interlengua que presentan rasgos de género y número diferentes a la lengua estándar. Se estudió el sintagma nominal núcleo (snn) y se consideraron algunos casos del sintagma nominal (SN). En cuanto al reconocimiento automático, se utilizó Smorph y Mps. La creación de dos tipos de reglas de reagrupamiento permitió que se extraigan tanto los SNN y Sn que coinciden con el sistema lingüístico del español, como los snn y los SN idiosincrásicos de la interlengua.

REFERENCIAS

[1] Méndez, B (2009) : “Análisis automático de la interlengua: asignación de género y número diferentes a la lengua estándar en el sintagma nominal núcleo(snn)” , en *La interlengua de los aprendientes del español como L2*, Grupo infosur, UNR, Rosario, ediciones Juglaria.

[2] Ambadiang, Théophile (1999): La flexión nominal. Género y número, en *Gramática descriptiva de la lengua española*, dirigida por Ignacio Bosque y Violeta Demonte. Espasa, Madrid.

[3] Rigau, Gemma (1999): “La estructura del sintagma nominal: los modificadores del nombre”, en *Gramática descriptiva de la lengua española*, dirigida por Ignacio Bosque y Violeta Demonte. Espasa, Madrid.

[4] Solana, Z y Rodrigo (2005), “El sintagma nominal núcleo”, en *Desarrollo, implementación y uso de modelos para el procesamiento automático de textos* (ed. Víctor Castel) Facultad de filosofía y letras, UNCUYO

[5] Ait-Mokhtar S.(1998) *L'analyse presyntaxique en une seule étape*. Tesis de Doctorado, GRIL, Univ.Blaise Pascal.

[6] SOLANA, Zulema (2006), *Lingüística Informática y enseñanza de lenguas*, Rosario, Ediciones Juglaría.