

FRANCESCA DALLE PEZZE ELISA SARTOR

EL VOCABOLARIO ITALIANO E SPAGNOLO (1620) DE LORENZO FRANCIOSINI: REPRESENTACIÓN DEL DISCURSO REPETIDO PARA SU DIGITALIZACIÓN

Università degli Studi di Verona

Resumen

La investigación examina los retos que plantea el etiquetado del discurso repetido en el *Vocabolario italiano e spagnolo* de Franciosini (1620) según el estándar TEI Lex0 empleado en el proyecto TELEI (*Tesoro Lexicográfico Español-Italiano*). Se trata de un estudio piloto que permite prever problemas de etiquetado, extensibles posteriormente al conjunto de la obra, que atesora un caudal de expresiones fijas que no es fácil sistematizar, debido a la incoherencia de su representación lexicográfica.

palabras clave: Franciosini, lexicografía bilingüe italiano-español, discurso repetido, codificación XML-TEI, siglo XVII

Abstract

Vocabolario italiano e spagnolo (1620) by Lorenzo Franciosini: representation of phrases for its digitalisation

This research examines the challenges posed by the tagging of phrases in Franciosini's Vocabolario italiano e spagnolo (1620) using the TEI Lex0 standard according to the TELEI (Teso Lexicográfico Español-Italiano) project protocol. This is a pilot study that allows the detections of tagging problems that could later be applied to the whole dictionary, which contains a wealth of fixed expressions that are not easy to systematize due to the inconsistency of their lexicographic representation.

keywords: Franciosini, Spanish-Italian bilingual lexicography, phrases, XML-TEI encoding, 17th Century

1. Introducción

El objetivo de este artículo es ofrecer una contribución al estudio del discurso repetido en el *Vocabolario italiano e spagnolo* de Lorenzo Franciosini, cuya primera edición se publicó en Roma en 1620¹, enfocada hacia las reflexiones y los retos que su volcado en formato XML-TEI supone desde un punto de vista teórico-metodológico.

En efecto, este trabajo se enmarca en el proyecto PRIN 2022 “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*”, (TELEI), cuyo objetivo es el de poner a disposición en una plataforma de acceso abierto el patrimonio lexicográfico bilingüe italiano-español en formato digital e hipertextual.

En el presente estudio vamos a analizar los retos que supone el tratamiento informático del *discurso repetido*. Elegimos este hiperónimo, de raíz coseriana, por dos razones principalmente: en primer lugar, porque, ante la asistematicidad de Franciosini, preferimos una etiqueta más abarcadora que diferenciadora; en segundo lugar, porque el lexicógrafo no se limitó a los subtipos que para el siglo XXI son “clásicos”, como refrán, locución, colocación, sino que lematizó otras estructuras, más periféricas, como unidades poliléxicas de la lengua común o de dominios especializados.

El enfoque de este estudio ha sido el de cómo tratar la fraseología en una plataforma digital, acotando los problemas metodológicos que surgen a la hora de volcar un contenido lexicográfico antiguo en un medio digital moderno. Puesto que se trata de un proceso *in fieri*, creemos que el análisis realizado sobre un porcentaje del archivo txt, fruto de la transcripción completa de la obra, no perjudica la validez de los resultados, ya que nuestro estudio piloto nos permite prever precisamente eventuales problemas de etiquetado, extensibles en un segundo momento a todo el conjunto de la obra. Dentro de las diferentes fases de trabajo este estudio piloto representa la ocasión para operar una reflexión global e idear soluciones operativas.

* Este trabajo se ha realizado en el marco del proyecto “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*”, financiado por la Unión Europea-Next Generation EU, Misione 4 Componente 1 CUP C53D23004010006 / PRIN 2022 MUR 20229W73WR. Aunque el presente ensayo se ha elaborado de forma conjunta, a efectos académicos Francesca Dalle Pezze es responsable de los apartados 1, 2 y 6 y Elisa Sartor de los apartados 3, 4 y 5.

1 El estudio se ha llevado a cabo sobre esta edición, que es la de referencia para nuestra digitalización.

2. Estado de la cuestión

A principios del siglo XXI y en una coyuntura de general interés por la labor de Lorenzo Franciosini, varios estudios coincidieron en tratar, entre otros, un aspecto específico —si bien sobresaliente— de su *Vocabolario*: la lematización de las expresiones pluriverbales. Así pues, Martínez Egido y Ruiz Gurillo (2004: 1055) destacaron la sorpresa que nos reserva no solamente descubrir su alto número sino también apreciar su variada riqueza. Ya anteriormente Scandola (2001) había resaltado la especial inclinación del autor hacia el mundo fraseológico, que se refleja en el caudal de expresiones recogidas en sus *Diálogos apazibles*. En esta obra —traducción de los *Pleasant and Delightfull Dialogues in Spanish and English* del lexicógrafo John Minsheu—, Franciosini muestra su comportamiento de hispanista no nativo hacia todos aquellos coloquialismos y construcciones que ni una gramática ni un diccionario podían recoger exhaustivamente, la primera por su planteamiento mismo, el segundo por falta de espacio (Scandola 2001: 82). Asimismo, la riqueza del material fraseológico del *Vocabolario* abrió el camino a la investigación sobre su relación con el universo conversacional contenido en los *Diálogos apazibles*, puesto que ambos tipos de obra persiguen una finalidad parecida, aunque no coincidente. Martínez Egido (2008), entonces, analizó los elementos fraseológicos repetidos en ambas obras y también los añadidos en los *Diálogos*, que se publicaron seis años después del repertorio.

Sin embargo, la falta de sistematicidad del autor en la inserción de unidades pluriverbales fue algo resaltado por todos los estudiosos (Alvar 2002: 59; Acero Durández 2004: 202; Martínez Egido, Ruiz Gurillo 2004: 1057), quienes señalaron su dependencia de una práctica lexicográfica todavía en desarrollo.

Los criterios de lematización de la fraseología cuentan con una bibliografía extensa, siendo uno de los problemas más llamativos de la técnica lexicográfica (Martínez Marín 1996; Ruiz Gurillo 2000; Porto Dapena 2002: 148-170). Se puede afirmar que el interés por el *Vocabolario* de principios del siglo XXI coincidió con la simultánea propensión de una parte de los estudios lingüísticos por la compleja cuestión de esos criterios. El estudio diacrónico del quehacer fraseográfico se convirtió en uno de los numerosos bancos de pruebas para la teoría general, que se interrogaba sobre cuánta fraseología insertar en un diccionario modélico, qué lugar reservarle (en la macroestructura o en la microestructura) y qué elemento del sintagma escoger para registrar la unidad (Acero Durández 2004: 199-200).

La cuestión del tratamiento reservado por parte de Franciosini a las expresiones pluriverbales revistió gran importancia porque, empezando por un enfoque simplemente cuantitativo, sirvió para realizar un cálculo global del tipo de entradas que recoge el autor, labor llevada a cabo por Martínez Egido (2002a: 183). Por lo tanto, en la primera parte el estudioso detectó el 9,5 % de sintagmas, como *palla di legno*, y el 6,67 % de oraciones, como *raffrenar le voglie* (2002a: 184), mientras que en la segunda comprobó que el 13,59 % de las entradas está representado por sintagmas, como *llaga reciente*, y el 11,97 % corresponde a oraciones, como *labrar la tierra* (2002a: 206). En el estudio, asimismo, se destacaba el acrecentamiento de estos dos tipos de entrada compleja en el paso de la primera a la segunda parte (2002a: 206), que puede cuantificarse en un 10 %. Descripciones exhaustivas como estas demostraron claramente la diferencia de composición de los dos volúmenes, independientes más que complementarios y con finalidades diferentes. Como señaló Alvar (2002: 53), la primera parte (italiano-español) se asemeja a un diccionario monolingüe italiano con la adición de equivalentes en español. Esta peculiar combinación de lo monolingüe con lo bilingüe demuestra la prontitud del autor en aprovechar los avances culturales contemporáneos, como la reciente publicación del *Tesoro de Covarrubias* (2002: 57).

En Acero Durández (2004) se estableció una minuciosa taxonomía de todos los tipos de unidades pluriverbales, basada en la clasificación de Corpas Pastor (1996). Según las calas de la autora, el lexicógrafo recogió profusamente unidades pertenecientes a las tres esferas, a saber, colocaciones, locuciones y enunciados fraseológicos. Se detectaron también algunas tendencias en la general falta de sistematicidad: las colocaciones son más frecuentemente entradas autónomas en la macroestructura (Acero Durández 2004: 204), mientras que los refranes y las fórmulas rutinarias forman parte habitualmente de la microestructura (2004: 207-8). Además, en cuanto a las locuciones, algunas se recogen incluso bajo un lema que no forma parte de la unidad pluriverbal, como *a horcajadas*, tras *horcador* (2004: 205).

Fragapane (2020) ha hecho hincapié en la presencia de elementos extrafraseológicos que constituyen el contorno sintáctico, que se convierte, sin embargo, en información muy valiosa para el usuario a la hora de reconstruir la forma citativa de la unidad misma: se trata de la explicitación de los actantes, mediante pronombres como *uno* o *algo* (*convenir a uno en juicio*) (170).

Otro asunto importante fue la detección de las (si bien escasas) notaciones metalingüísticas que Franciosini emplea para indicarle al usuario que se encuentra frente a un elemento del discurso repetido. Así pues, Martínez Egido y Ruiz

Gurillo (2004: 1058) señalaron *modo di dire*, que efectivamente se asocia en numerosas ocasiones a locuciones verbales —que son la categoría más nutrida de las recopiladas por el lexicógrafo, según el recuento de Acero Durández (2004: 205)—, pero también a locuciones adverbiales, a enunciados fraseológicos e incluso a interjecciones. Por lo tanto, el tratamiento de las unidades fraseológicas en la codificación XML del diccionario de Franciosini requiere una definición más bien ‘operativa’ de UF, lo que se pretende hacer en el apartado 3. Finalmente, desde la vertiente de la lexicografía del portugués, llega una fructuosa aportación al estudio de la sistematización de la fraseología en la codificación de un diccionario en XML (Tasovac, Salgado, Costa 2020), que orienta también nuestras reflexiones en el apartado 5.

3. Metodología

La conversión en formato txt de los dos volúmenes del diccionario nos ha brindado la oportunidad de realizar búsquedas en el texto por palabras clave. Esto nos ha permitido encontrar no solo las expresiones pluriverbales lematizadas como lemas o sublemas, sino también aquellas que se hallan en la microestructura, así como las marcas lexicográficas que las acompañan. A este respecto, cabe aclarar que, pese a que la transcripción de ambos tomos está completa, la investigación se ha llevado a cabo empleando solo las páginas revisadas de forma definitiva a la fecha de redacción del presente estudio (enero de 2025), con el objetivo de obtener datos más fiables desde el punto de vista textual. De ahí que se haya analizado el 46,7 % de las entradas que conforman el primer volumen y el 35,3 % de las del segundo, con una composición variable: algunas letras se han considerado parcialmente, otras en su totalidad (E, F, H, L, N, O, Q, U, V, X, Z del primer tomo y G, L, N, Ñ, O, Q, X, Z del segundo); finalmente, algunas han quedado excluidas por no estar terminado el proceso de revisión textual, concretamente, D, G, T de la parte italiano-español y F, S, T, U, V de la parte español-italiano.

Un rastreo efectuado a través de las palabras clave “modo di dire” y “proverbio” ha permitido extraer aquellos fraseologismos que el autor consideraba útil señalar a los usuarios de su diccionario; en concreto se han encontrado 80 unidades pluriverbales marcadas con la etiqueta “modo di dire” y 8 con “proverbio”. Citamos como ejemplos, respectivamente,

(1) moliente, y corriente. [è un modo di dire che significa che uno è apparecchiato, e lesto a far ogni cosa.

(2) quando vieres la barba de tu vezino pelar, echa la tuyu in remojo. [quando la casa del tuo vicino abbrucia, porta l'acqua a casa tua. Inferisce questo proverbio, che quando vediamo intervenir qualche disgrazia a persona della nostra condizione, e sorte, haviamo a pensar che un'altro giorno toccherà anche a noi.

Otras pesquisas realizadas con la palabra clave “suol” han puesto de relieve 18 unidades pluriverbales que muestran una fuerte solidaridad semántica entre sus componentes, encajando con la actual definición de colocación, por ejemplo,

(3) legione, o gran numero, come si suol tal volta dire de'demoni. [legion de demonios.

(4) manàr, agua o otra cosa liquida. [scaturire, o surgere acqua, o altra cosa liquida, e si suol dire comunemente delle fonti, che scaturiscon acqua.

Desde luego, este primer sondeo ha dejado claro que Franciosini no emplea las notaciones metalingüísticas en modo sistemático, como ya destacaron Martínez Egido y Ruiz Gurillo (2004). Por un lado, parece atribuir las etiquetas “modo de dire” y “proverbio” a las unidades que actualmente consideramos locuciones y enunciados fraseológicos², respectivamente, mientras que la fórmula “suol” por lo general introduce sintagmas caracterizados por solidaridad semántica, que adscribiríamos al ámbito de las colocaciones. Por otro lado, este patrón no se aplica de forma rigurosa —véase, por ejemplo, el enunciado

(5) buen jubon me tengo en Francia. [è un modo di dire che si usa quando uno ci promette una cosa, che al presente non l'ha, ne ce la puol dare.

y, además, cuantiosos fraseologismos no llevan ningún tipo de marcación que los describa, como

(6) perro ladrador, poco mordedor. [can che abbaia non morde.

Por lo tanto, se ha decidido construir un corpus piloto³ para comprobar en qué

2 Nos apoyamos en la clasificación tradicional de Corpus Pastor (1996) por ser bien asentada y aceptada en el contexto de la fraseología actual del español, además de suficientemente abarcadora en relación con las cuestiones metodológicas que nos atañen.

3 En estudios anteriores dedicados al tratamiento de la fraseología en Franciosini, se han analizado corpus compilados a partir de criterios distintos. Por ejemplo, Martínez Egido (2002b) selecciona una muestra de entradas (263) procedentes de la primera parte del diccionario (italiano-español), concretamente de las letras B, L y T. El corpus del trabajo de Ruiz Gurillo y Martínez Egido (2006),

medida Franciosini recoge unidades fraseológicas en la nomenclatura de su diccionario y las emplea en la microestructura de las entradas, con vistas a examinar el tratamiento lexicográfico de las expresiones pluriverbales, incluso las que él no marca como tales. En este trabajo se ha decidido emplear una definición muy amplia de *expresión pluriverbal* para poder abarcar distintas tipologías de discurso repetido, hasta las colocaciones, que representan una categoría resbaladiza, debido a la variabilidad de la solidaridad semántica de los elementos que las componen, especialmente en perspectiva diacrónica. Además, estimamos útil para este estudio tener en cuenta también las unidades poliléxicas nominales con sentido estable y significado tanto (semi)composicional como no composicional, que Montoro del Arco (2008) denomina *formaciones nominales pluriverbales*: esto nos permite considerar los compuestos sintagmáticos de la lengua común así como los que pertenecen al lenguaje especializado (por ejemplo, al dominio de la botánica), que de otra forma quedarían excluidos de nuestro análisis —y posterior tratamiento informático— por no ser fraseologismos *strictu sensu*.

El corpus se ha construido a partir de una selección de somatismos, puesto que, como ya se ha demostrado en numerosos estudios (Navarro 2008, Mellado Blanco 2023, entre otros), estos son especialmente productivos; concretamente, en esta ocasión la búsqueda se ha realizado empleando lexemas relacionados con la cabeza y la cara: *barba, ojo, diente, oreja, nariz y boca*⁴, sus equivalentes en italiano y las formas del plural en ambos idiomas. La búsqueda ha devuelto 17 expresiones pluriverbales que contienen la palabra *barba(s)/barb**, 58 con *ojo(s)/occhi(o)*, 15 con *diente(s)/dent**, 11 con *oreja(s)/orecch**, 16 con *nariz(es)/nas** y 31 con *boca(s)/bocc**, que se estudiarán desde el punto de vista cualitativo en el apartado siguiente.

Para ello, se han consultado dos obras lexicográficas contemporáneas que Franciosini tuvo muy presentes a la hora de componer su diccionario bilingüe, a saber, el *Vocabolario degli Accademici della Crusca* (1612) y el *Tesoro de la lengua castellana o española* de Covarrubias (1611), que fueron una fuente directa para la elaboración de la microestructura del tomo italiano-español y español-italiano, respectivamente (Martínez Egido 2002b y 2003).

en cambio, se ha construido a partir de cuatro verbos (*dar, echar, saber y sacar*), con el resultado de estar formado únicamente por locuciones verbales.

4 La creación de un corpus piloto a partir de somatismos derivados de partes móviles del cuerpo humano se ha llevado a cabo recientemente en otro estudio fraseográfico basado en un diccionario del s. XX, a saber, el *Diccionario de uso del español* de María Moliner. Para profundizar en la metodología empleada y los resultados del estudio véase Dal Maso y Sartor (2025).

4. Resultados

El análisis cualitativo del corpus de estudio nos ha permitido mirar desde cerca los procedimientos lexicográficos adoptados por Franciosini, comprobando algunos datos señalados en estudios anteriores y que, sin embargo, no exploraremos por exceder el propósito de este trabajo. Baste con citar la marcación lexicográfica, aspecto ya estudiado por Acero Durández (2002), de la que se encuentra un ejemplo múltiple en la entrada

(7) hallar horma de su capato. [per parlare chiaramente, e secondo l'uso, e a proprietà di questo, si dice correntemente, se bene con poca honesta, trovar naso da suo'culo, cioè trovar persona che ti sappia rispondere, e ti mostri il viso.

Franciosini, como se puede ver, proporciona información sobre la frecuencia de uso (*si dice correntemente*) y a la vez advierte sobre la malsanancia del fraseologismo italiano (*con poca honesta*).

Desde el punto de vista léxico, merece la pena observar las estrategias empleadas para establecer equivalencias semánticas entre los dos sistemas lingüísticos lo más adecuadas posible, básicamente a través de dos procedimientos, a saber: una expresión pluriverbal equivalente en la lengua meta, aunque formalmente distinta, o bien una explicación del significado de la unidad. Estas dos técnicas a veces se emplean de forma híbrida, proporcionando un equivalente y también una glossa, especialmente cuando la lengua de redacción es el italiano, tanto si es la lengua fuente como si es la lengua meta. En los ejemplos siguientes se muestran casos que caben en la primera tipología (8-12), en la segunda (13-14), o que presentan una mezcla de las dos (15-18):

- (8) ojos que no ven, coraçon no quebrantan. [quello che gli occhi non veggano, il cuor non crede. (*s.v. ojo*)
- (9) batir los dientes. [battere i denti. (*s.v. batir los dientes*)
- (10) boca de leon. [leonina herba. (*s.v. leon*)
- (11) nuve del ojo. [maglia dell'occhio. (*s.v. nuve del ojo*)
- (12) lienço de narizes. [fazzoletto. pezzuola. (*s.v. lienço*)
- (13) a poca barba, poca verguença. [inferisce che i giovani con la sua poca prudenza soglion esser talvolta molto sfacciati, e persuntuosi. (*s.v. barba* ES)
- (14) muelas de gallo. [si dice a colui che non ha nissun dente. (*s.v. gallo* ES)
- (15) levarsi le mosche d'intorno al naso, vale non si lasciar burlare ne far ingiuria. [no

sufrir cosquillas. (*s.v. mosca* IT)

(16) a bulto. [a occhio, senza pesare o misurare. (*s.v. a bulto*)

(17) diente de ajo. [spicchio d'aglio, cioe una di quelle particelle in che si divide. (*s.v. cabeça de ajo*)

(18) morselletto, boccone di pasta di varie sorte d'ingredienti, che si fanno da speziali. [tabletta, o pastilla de boca. (*s.v. morselletto*).

Desde un punto de vista taxonómico actual, las unidades pluriverbales citadas pertenecen a distintas categorías, a pesar de que, en ninguno de los ejemplos mencionados, Franciosini emplee etiquetas que lo señalen: se podrían considerar enunciados fraseológicos los ejemplos (8) y (13), locuciones (14), (15) y (16)—respectivamente nominal, verbal y adverbial—, colocaciones (9) y (17), compuestos *lienço de narizes* y *pastilla de boca* y, finalmente, formaciones poliléxicas propias del lenguaje de especialidad *boca de leon* (del dominio de la botánica) y *nuve del ojo* (término de la medicina).

Otra cuestión que merece la pena destacar es la posición de la unidad pluriverbal en la entrada o subentrada. Se han encontrado casos en los que la expresión pluriverbal es el lema mismo (véase *infra*, 19-20) o constituye un sinónimo del lema (*a occhio e croce, orejas de Abad*, 21 y 22, respectivamente); en otros casos se halla en la microestructura y puede ser el equivalente del lema (23 y 24, donde se ofrecen hasta tres variantes fraseológicas: *guardar con mal occhio uno, non gli far buon viso, [non gli far] buona cera*) o bien un sinónimo del equivalente (*digrignare i denti y guarda la gamba*, 25 y 26, respectivamente):

(19) vale un ojo de la cara. [modo d'esaggerare il prezzo d'una cosa, vale un'occhio. (*s.v. ojo*)

(20) dexelo, que no le hiede la boca. [modo di dire che usano le bottegaie quando non gli si dà per la lor mercanzia quanto domandano. (*s.v. heder*)

(21) a vista, avverbialmente giudicare a vista, cioè con la semplice vista, senza venire ad altro cimento, che diremmo anche giudicare a occhio, e croce. [a ojo. a bulto. (*s.v. a vista* IT)]

(22) lasañas, o orejas de Abad. [certa pasta fritta nella padella in forma di frittelle, ò cresPELLI. (*s.v. lasañas*)

(23) mangiar a mal in corpo, cioè mal volentieri, come tal volta fanno gli ammalati. [comer a regaña dientes. (*s.v. mangiar a mal in corpo*)

(24) estrañarse. [si dice il guardar con mal occhio uno, e non gli far buon viso, o buona cera. (*s.v. estrañarse*)

(25) chillar. [stridere, e far rumore come quando si frigge, o si strofinano i denti, l'uno con l'altro, che dichiamo propriamente sdirugginare i denti, o dignignare i denti. (*s.v. chillar*)

(26) abre el ojo. [maniera di dire per avvertire. guarda la gamba. (*s.v. ojo*).]

Es oportuno señalar, sin embargo, que se dan casos ambiguos en los que resulta muy difícil averiguar en diacronía si un determinado sintagma constituía una forma de discurso repetido en la época de Franciosini, como en el ejemplo siguiente:

(27) A, in vece d'inguisa, ò come. Hâ i denti fatti à bischeri. [Tiene los dientes como clavijas, o a manera, o a semejança de clavijas. (*s.v. a IT*).]

Concretamente, *denti fatti a bischeri* se recoge como ejemplo en una entrada del diccionario de los académicos de la Crusca (1612). Se trata de una cita de una *novella* de Boccaccio, que representó uno de los modelos, junto con Petrarca, para la codificación de la lengua italiana culta en las etapas humanista y renacentista: “E mi par vederti morderle, con cotesti tuoi denti fatti a bíscheri, quella sua bocca vermigliuzza, ec.” (*s.v. bischero*). Esta misma cita se mantiene hasta la cuarta edición del *Vocabolario* (1729-1738) y, en forma abreviada, se emplea también para ejemplificar el lema *mordere* de la primera a la cuarta edición. Debido a la falta de documentación de este sintagma en sincronía, podemos suponer que se usó como ejemplo en las obras lexicográficas del s. XVII y posteriores⁵ por el prestigio literario de su autor, sin pasar a convertirse en un elemento de discurso repetido.

Otro caso límite es la ocurrencia, generalmente en la microestructura, de enunciados que tienen una estructura que trae a la memoria los dichos proverbiales o las sentencias y que, sin embargo, no se documentan en otras fuentes consultadas⁶:

(28) A uno, che parlava molto, e dava poco, disse una gentil donna il miglior huomo del mondo saresti, se le serrature che havete nella borsa, l'havessi nella bocca. [A uno, que hablava mucho, y dava poco, le dixo una Señora. El mejor hombre del mundo seriades, si los cerraderos que teneis en la bolsa, tuviessedes en la boca. (*s.v. borsa*).]

⁵ Señalamos que la misma cita, en una versión más extensa y con la marca de uso *per similitudine*, se recoge también en el *Grande dizionario della lingua italiana* de Salvatore Battaglia (1961-2004), *s.v. bischero*.

⁶ Como es sabido (Alvar 2002: 43), Franciosini conocía y se sirvió de las breves historietas registradas en la *Floresta española* de Melchor de Santa Cruz (1.^a edición, 1574), como la siguiente: “A Vno que tenia vna cuchillada por la cara, y hablaua demasiadamente, dixo vno, Pues teneys rienda en la cara, tened freno en la boca” (Santa Cruz, 1576: 238v).

Desde luego, no se considera que sean formas de discurso repetido sino más bien glosas o breves anécdotas con las que Franciosini pretende ilustrar ejemplos de uso en las (sub)entradas.

A este propósito, finalmente, Fragapane (2020: 175) subraya el hecho de que la finalidad didáctica, que prima en la obra sobre las demás, acaba convirtiendo el límite entre fraseología y ejemplos de uso bastante borroso, ya que varios casos se ubican en una vía intermedia entre lo fraseológico y lo no fraseológico. Para ilustrar este fenómeno la autora menciona los ejemplos de *esto limpia es, bien se puede comer sin asco y yo no puedo comer de asco*, bloques que, más que verdadero discurso repetido (ya fijado e institucionalizado), son fórmulas que el usuario podría proyectar en su propio acto comunicativo si la situación lo permitiera.

5. Etiquetado con XLM-TEI Lex0

La pregunta de investigación que nos planteamos, no solo teórica sino práctica también por las implicaciones a la hora de volcar las entradas y subentradas en la plataforma XLM-TEI, atañe al tipo de información sobre las unidades pluriverbales que el usuario querrá o podrá obtener cuando consulte el diccionario de Franciosini en formato digital. El estándar Lex0 empleado en el proyecto TELEI (*Tesoro Lexicográfico Español-Italiano*) nos brinda la oportunidad de etiquetar las unidades pluriverbales con *multiword expression (mwe)*, lo que va a permitir recuperar la información a la hora de realizar consultas y de emplear la etiqueta *mwe* como filtro.

En el lenguaje de codificación empleado en el proyecto TELEI, las líneas de código que sirven para identificar una unidad pluriverbal en la macroestructura seguirán el modelo siguiente, basado en el ejemplo *nuve del ojo*:

```
<entry type="mainEntry" xml:id="Francio1620.nuve del ojo." xml:lang="es">
  <form type="lemma">
    <orth>nuve del ojo</orth></form>
    <gramGrp><gram type="mwe"/>
```

En concreto, la primera línea sirve para atribuir un *id* (identificativo) al lema y clasificar la expresión pluriverbal como perteneciente a la primera edición del diccionario de Franciosini, especificando la lengua del lema, en este caso el español⁷;

⁷ No hay que olvidar que la plataforma TELEI almacenará en formato XML todas las obras

la segunda línea caracteriza la unidad como lema del diccionario, mientras que la etiqueta *<orth>* es la que permite ingresar manualmente la expresión pluriverbal (en nuestro ejemplo, *nuve del ojo*); finalmente, la última línea permite la caracterización gramatical del lema como *multiword expression*.

Ahora bien, a fin de que el usuario pueda aprovechar plenamente los avances en la consulta que la digitalización supone, es necesario que todas las unidades pluriverbales sean correctamente etiquetadas para ser luego recuperadas a través del filtro *mwe*. Desde el punto de vista operativo, esto conlleva marcar con la etiqueta *mwe* todos los fraseologismos que el mismo Franciosini distinguió a través de las etiquetas “proverbio”, “modo di dire”, “si suol dire”, etc., tanto los que siguen vigentes (por ejemplo, *hasta los ojos*) como los que están desusados en la época actual (*no me dixo negro has el ojo*). Además, es preciso etiquetar también los demás elementos que se reconozcan como formas de discurso repetido en diacronía y sincronía (por ejemplo, la variante ortográfica *a regaña dientes*, hoy en desuso), incluso aquellos que hayan quedado obsoletos (*hombre de barba*), considerando también las unidades pluriverbales de uso general (*ojito del aguja*) o propias del lenguaje de especialidad (*oreja de ratón*) que caben bajo el marbete de expresión pluriverbal, esto es, *multiword expression*. Obviamente, los ejemplos (27) y (28), discutidos en el apartado anterior, quedarían excluidos por no ser formas de discurso repetido; concretamente, el caso (28) se etiquetaría como *gloss*, o sea, glosa.

Coherentemente con los criterios compartidos que rigen la creación de la plataforma TELEI, las expresiones pluriverbales que se clasifiquen como *multiword expression* se etiquetarán como *entry* (lema) si aparecen en la macroestructura; si conforman una subentrada, en cambio, se etiquetarán como una *entry* anidada en la etiqueta *<sense>*, que corresponde al sentido o acepción. Las ocurrencias en la microestructura, a su vez, se clasificarán como *multiword expression* dentro del *sense* de la entrada correspondiente. Esta jerarquización de la información se debe, por un lado, a las características lexicográficas del diccionario mismo, que registra expresiones pluriverbales tanto en la macro- como en la microestructura, según se ha comentado en el apartado anterior, y, por otro, a la estructura del estándar Lex0.

De acuerdo con Tasovac, Salgado y Costa (2020: 51), “TEI Lex0 is fully capable of consistently marking up poly lexical units as constituent parts of the

lexicográficas digitalizadas en el marco del proyecto “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*”, por lo tanto, es imprescindible que cada elemento esté vinculado en forma unívoca con el diccionario correspondiente.

dictionary macro- and microstructure, regardless of whether they appear as head-words in independent entries, or in nested entry-like structures inside entries for monolexical units". Desde luego, a diferencia del estudio de caso presentado por estos autores, que trabajan en la codificación de un diccionario actual de la lengua portuguesa, para la codificación de Franciosini se ha decidido desestimar algunas de las funciones de Lex0 que permite, entre otras cosas, marcar el núcleo de la expresión poliléxica o señalar los compuestos con guion (*compounds*), atendiendo a los rasgos propios de una obra lexicográfica del s. XVII.

6. Conclusiones

El proyecto PRIN 2022 “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*” (TELEI) y, más específicamente, la digitalización del *Vocabolario italiano e spagnolo* de Lorenzo Franciosini —obra cumbre de la producción de diccionarios español-italiano durante el siglo XVII— nos pone ante el reto de un adecuado tratamiento del discurso repetido.

El *Vocabolario italiano e spagnolo* de Lorenzo Franciosini sigue brindando hoy en día ocasiones de interés, no solamente cultural (por el rico y abigarrado caudal de informaciones que ofrece), sino también para la lexicografía informatizada y la filología actuales. El rastreo de los elementos que forman parte del discurso repetido, tanto si es prototípico como periférico, nos depara numerosos desafíos a la hora de posibilitar una búsqueda provechosa por parte del usuario de la plataforma digital e hipertextual.

Como hemos intentado demostrar, la reflexión teórico-metodológica que hemos llevado a cabo tendrá un impacto en la codificación; sin embargo, al ser este un trabajo en progreso, los posibles problemas y desajustes deberán ser tomados en consideración como retroalimentación para mejorar el proceso. Por lo tanto, sigue abierta una vía para futuras investigaciones y mejoras.

Bibliografía citada

- Accademiici della Crusca (1612), *Vocabolario degli Accademiici della Crusca*, Venezia, Giovanni Alberti [15/01/2025] <<http://www.lessicografia.it/>>.
- Acero Durández, Isabel (2004), “El tratamiento de la fraseología española e italiana en un diccionario bilingüe del siglo XVII”, *Nuevas aportaciones a la historiografía lingüística. Actas del IV Congreso Internacional de la SEHL. La Laguna (Tenerife), 22 al 25 de octubre de 2003*, eds. Cristóbal José Corrales Zumbado; Josefa Dorta Luis; Antonia Nelsi Torres González; Dolores Corbella Díaz; Francisca del Mar Plaza Picón. Madrid, Arco/Libros: I, 199-210.
- Alvar Ezquerro, Manuel (2002), “El *Vocabolario italiano e spagnolo* de Lorenzo Franciosini”, *Lengua y diccionarios. Estudios ofrecidos a Manuel Seco*, eds. Pedro Álvarez de Miranda; José Polo Polo. Madrid, Arco/Libros: 37-61.
- Battaglia, Salvatore (1961-2004), *Grande dizionario della lingua italiana*, Torino, UTET [15/01/2025] <http://www.gdli.it>.
- Covarrubias, Sebastián de (1611), *Tesoro de la lengua castellana o española*, Madrid, Luis Sánchez [15/01/2025] <https://apps2.rae.es/ntlle/SrvltGUILoginNtlle>.
- Dal Maso, Elena; Sartor, Elisa (2025), “Fraseologismos somáticos en el DUE de María Moliner: un estudio desde la perspectiva de los aprendices de ELE”, *Revista Internacional de Lenguas Extranjeras*, 24: 1-23 [13/06/2025] <https://www.revistes.urv.cat/index.php/rile/article/view/3973/4716>.
- Fragapane, Federica (2020), “El tratamiento de las unidades fraseológicas en el diccionario de Franciosini”, *Phrasis. Rivista di studi fraseologici e paremiologici*: 167-178 [24/01/2025] <https://www.phrasis.it/rivista/index.php/rp/article/view/64>.
- Franciosini, Lorenzo (1620), *Vocabolario italiano e spagnolo*, 1.^a edición, Roma: Gio. Paolo Profilio, 2 vols.
- Martínez Egido, José Joaquín (2002a), *La obra lexicográfica de Lorenzo Franciosini: Vocabulario italiano-español, español-italiano (1620)*, Alicante, Biblioteca Virtual Miguel de Cervantes [21/01/2025] <https://www.cervantesvirtual.com/obra/la-obra-lexicografica-de-lorenzo-franciosini-vocabulario-italianoespanol-espanolitaliano-1620--0/>.
- Martínez Egido, José Joaquín (2002b), “Las fuentes de la microestructura en la primera parte del *Vocabulario italiano-español, español-italiano* (1620) de Lorenzo Franciosini”, *Estudios de lingüística*, 16: 5-35.
- Martínez Egido, José Joaquín (2003), “La utilización de fuentes en la realización de diccionarios: ¿Covarrubias (1611), fuente directa en la microestructura de Franciosini (1620)?”, *ELUA*, 17: 413-26.
- Martínez Egido, José Joaquín; Ruiz Gurillo, Leonor (2004), “La integración de la fraseología en los primeros diccionarios del español: el *Vocabulario* (1620) de Lorenzo Franciosini”, *Nuevas aportaciones a la historiografía lingüística. Actas del IV Congreso Internacional de la SEHL. La Laguna (Tenerife), 22 al 25 de octubre de 2003*, eds.

Cristóbal José Corrales Zumbado; Josefa Dorta Luis; Antonia Nelsi Torres González; Dolores Corbella Díaz; Francisca del Mar Plaza Picón. Madrid, Arco/Libros: II, 1055-1066.

Martínez Egido, José Joaquín; Ruiz Gurillo, Leonor (2006), “Las unidades fraseológicas en el inicio de la lexicografía del español (1495-1620)”, *Actas del VI Congreso Internacional de Historia de la Lengua*, eds. José Luis Girón Alconchel; José Jesús de Bustos Tovar. Madrid, Arco/ Libros: 1-13 [21/01/2025] https://rua.ua.es/dspace/bitstream/10045/24048/1/2006_Leonor_Ruiz_Congreso_Historia_Lengua.pdf.

Martínez Egido, José Joaquín (2008), “El español conversacional y su reflejo en los diccionarios de Oudin y Franciosini”, *El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica*, (eds.) Dolores Azorín Fernández; Belén Alvarado Ortega; Jaume Climent de Benito; M. Isabel Guardiola i Savall; Ruth M. Lavale Ortiz; Carmen Marimón Llorca; José Joaquín Martínez Egido; Xose A. Padilla García; Herminia Provencio Garrigós; Isabel Santamaría-Pérez; Larissa Timofeeva Timofeev; Elena Toro Lillo. Alicante, Biblioteca Virtual Miguel de Cervantes: 144-149 [21/01/2025] <https://www.cervantesvirtual.com/obra/el-espaol-conversacional-y-su-reflejo-en-los-diccionarios-de-oudin-y-franciosini-0/>.

Martínez Marín, Juan (1996), *Estudios de fraseología española*, Málaga, Librería Ágora.

Mellado Blanco, Carmen (2023), “Los somatismos y su transparencia semántica: los rasgos tipológicos”, *EntreLíneas*, 6: 1-23 [17/01/2025] https://entrelineasrevista.wordpress.com/wp-content/uploads/2023/03/monografico_entrelineas.pdf.

Montoro del Arco, Esteban T. (2008), “Relaciones entre Morfología y Fraseología: las formaciones nominales pluriverbales”, *Neologismo y morfología*, eds. Ramón Almela Pérez; Esteban Tomás Montoro del Arco. Murcia, Universidad de Murcia: 121-146.

Navarro, Carmen (2008), *Aspectos de fraseología contrastiva*, Verona, Fiorini Editore.

Porto Dapena, José-Álvaro (2002), *Manual de técnica lexicográfica*, Madrid, Arco/Libros.

Ruiz Gurillo, Leonor (2000), “Cómo integrar la fraseología en los diccionarios monolingües”, *Las lenguas de Europa: estudios de fraseología, fraseografía y traducción*, (coord.) Gloria Corpas Pastor. Granada, Comares: 261-274.

Santa Cruz, Melchor de (1576), *Floresta española de apothegmas o Sentencias*, Alcalá, Juan Íñiguez de Lequerica.

Scandola, Viviana (2001), “Lorenzo Franciosini: un interés declarado por la fraseología”, *Paremia*, 10: 79-84 [16/01/2025] https://cvc.cervantes.es/lengua/paremia/pdf/010/009_scandola.pdf.

Tasovac, Toma; Salgado, Ana; Costa, Rute (2020), “Encoding polylexical units with TEI Lex-0: A case study”, *Slovenščina 2.0*, 8(2): 28-57.

TEI (Text Encoding Initiative) [22/01/2025] <https://tei-c.org>.

TEI Lex-0 [22/01/2025] <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>.

Francesca Dalle Pezze es profesora titular de Lengua, Traducción y Lingüística Españolas en la Universidad de Verona, donde participa en el grupo de investigación *El léxico del español en su historia (LEHist)* y en el PRIN 2022 “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*”. Sus principales líneas de investigación se centran en la lexicografía, la lexicología y la historia de la lengua española.
francesca.dallepezze@univr.it

Elisa Sartor es profesora titular de Lengua, Traducción y Lingüística Españolas en la Universidad de Verona, donde participa en el grupo de investigación *El léxico del español en su historia (LEHist)* y en el PRIN 2022 “Un nuovo ambiente digitale per il recupero del patrimonio lessicografico: il *Tesoro digitale della lessicografia bilingue spagnolo-italiano*”, entre otros proyectos. Sus líneas de investigación se centran en el estudio histórico del léxico y la lingüística de corpus.
elisa.sartor@univr.it