



CLASIFICACIÓN SEMÁNTICA VERBAL EN TEXTOS ESCRITOS DIRIGIDOS A NIÑOS LECTORES DEBUTANTES MEXICANOS

ANA LUISA FORZÁN DE LACHICA/ MÓNICA ALVARADO CASTELLANOS/ MARTHA BEATRIZ SOTO MARTÍNEZ

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
ana.luisa.forzan@uaq.mx
moncalvar@gmail.com
moncalvar@gmail.com

RESUMEN

El objetivo de la presente investigación fue indagar cuáles y cómo son los verbos (clases semánticas) más frecuentes en los textos de alta divulgación proveídos a los niños mexicanos en momentos iniciales de alfabetización inicial. Con base en la metodología lexicográfica y la lingüística de corpus, se identificó el vocabulario fundamental de los verbos que estuvo integrado por 155 lemas, todos ellos extraídos de un corpus de 604 textos literarios, contenidos en los libros de lectura de la Secretaría de Educación Pública (para 1º y 2º de primaria). La clasificación semántica verbal de base que adoptamos fue la del proyecto de Alternancia de Diátesis y Esquemas Sintáctico – Semánticos del Español (ADESSE). Se encontró que de los seis tipos de procesos que según Halliday y Matthiessen (2004, 2006) son las representaciones o figuras gramaticales de un tipo de experiencia, el mayor porcentaje de ocurrencias se concentró en el proceso material (42.09%) y el proceso mental (23.33%) que en conjunto representan el 64.42% del total del vocabulario fundamental. Las concentraciones mayores, materiales y mentales, responden a la categoría gramatical de los procesos que expresan las experiencias del mundo exterior y del mundo de la conciencia respectivamente. El presente trabajo es de naturaleza instrumental, de manera que su utilidad está relacionada con decisiones metodológicas para futuros estudios psicolingüísticos, evaluación del lenguaje y decisiones didácticas que servirán de base para favorecer la el aprendizaje de la lengua.

Palabras clave: Verbos escritos, análisis de corpus, vocabulario fundamental, clases semánticas





INTRODUCCIÓN

El objetivo de la presente investigación fue indagar cuáles y cómo son los verbos (clases semánticas) más frecuentes en los textos previstos para los niños mexicanos en momentos iniciales de alfabetización. Forma parte de un proyecto mayor cuyo propósito fue describir el vocabulario fundamental en los textos escritos para niños mexicanos. Del que ya se concluyó el vocabulario de sustantivos y esta entrega da cuenta de los verbos, específicamente de sus clases semánticas. El proyecto es coordinado por la Dra. Mónica Alvarado Castellanos.

El problema de investigación responde a un planteamiento de la lingüística aplicada asociada al desarrollo de un recurso instrumental descriptivo de la lengua escrita a la que están expuestos los niños mexicanos recién alfabetizados que cursan regularmente los grados de 1º y 2º de primaria. La lingüística y la psicología no pueden trabajar al margen de las descripciones de la lengua oral y escrita. Por lo que, las descripciones de este estudio eventualmente pueden servir para realizar estudios cognitivos (psicolingüísticos) sobre todo en lo referente al reconocimiento de palabra, lectura de palabras en contexto, comprensión y fluidez lectora. Adicionalmente, desde la psicología de la educación, pueden ser útiles para la toma de decisiones relacionadas con la adquisición de la lengua.

En consecuencia, nuestro estudio pretendió contestar a dos cuestiones centrales:

- *¿Cuáles y cómo son los verbos más frecuentes en los textos escritos para niños mexicanos en momentos iniciales de alfabetización?*
- y más específicamente,
- *¿cuáles son las clases semánticas más frecuentes de los verbos que componen el vocabulario fundamental del C-TEN?*

Para contestar a las preguntas de investigación nos planteamos las siguientes tareas:

a) Identificación del vocabulario fundamental. El primer procedimiento (Halliday, 2004b; Lara, et al, 1979; Sinclair, 2005) para la identificación del vocabulario fundamental fue lematizar los vocablos del corpus de verbos. Sobre el conjunto de verbos, se estableció la forma infinitiva de los mismos como cabeza léxica o lema que enlistamos de mayor a menor frecuencia.





Para determinar el vocabulario fundamental del corpus utilizamos el de frecuencia de aparición igual o mayor a 20 (F \geq) establecido por *Centre de Recherche et d' Etude pour la Diffusion du Français* utilizado para definir el vocabulario fundamental del francés.

Establecimos una cláusula de exclusión para el núcleo léxico o vocabulario fundamental con el que trabajamos en el análisis sintáctico-semántico aunado al criterio de frecuencia de uso. El criterio de exclusión que consideramos fue no incluir los verbos ya identificados en el vocabulario fundamental que no tienen significación plena como son los copulativos y los de soporte o de apoyo.

b) Procedimiento para la clasificación semántica. Las clasificaciones semánticas verbales son múltiples y variadas. La que utilizamos en este estudio fue la desarrollada en el Proyecto de Alternancia de Diátesis y Esquema Sintáctico – Semántico del Español (ADESSE) (García-Miguel, Costas y Martínez, 2003). El proyecto ADESSE que actualmente se lleva a cabo en la Universidad de Vigo, consiste en la creación de una base de datos con información sintáctica – semántica de los verbos del español. Para su realización parten a su vez de los datos de la Base de Datos Sintácticos del Español Actual (BDS). Elegimos esta clasificación porque asigna la clasificación semántica partiendo de los contextos de uso. Lo que nos pareció relevante para reflexionar sobre los verbos a los que están expuestos los niños en los momentos iniciales de alfabetización dado que no se restringen a un significado léxico únicamente sino también al construccional.

El fundamento teórico y la estructura de la clasificación de ADESSE consta de:

- El fundamento teórico se basa en Dixon (1991), Faber y Mairal (1999), FrameNet, Halliday (1985), Halliday y Matthiessen (2004, 2006), Levin (1993), Vázquez (2000), y WordNet.
- La jerarquía se presenta en un nivel de abstracción más general que son los seis tipos de procesos o macroclases (Halliday, 1985; Halliday y Matthiessen, 2004, 2006), 14 dominios conceptuales o clases (Faber et al., 1999; primera versión FrameNet; Dixon, 1991; WordNet: Fellbaum, 1998) y 52 marcos conceptuales, frames o subclases (Levín, 1993; Vázquez, et al., 2000).

Las tareas que realizamos para la clasificación fueron:





- reconocimiento y asignación de clases y subclases;
- identificación de la macro-clase; y
- agrupamiento de los niveles de concentración de las clases semánticas.

RESULTADOS

a) Resultados de la identificación del vocabulario fundamental

Antes de presentar los resultados es necesario describir el Corpus de Textos Escritos para Niños conformado por los libros de textos de lectura para niños de 1º y 2º de primaria editados por la Secretaría de Educación Pública bajo la consideración de que representan uno de los mayores tirajes editoriales de materiales escritos en México (en promedio 1, 700,000 cuando los tirajes comunes son de 100,000 ejemplares). Además, que son entregados al 100% de los niños que cursan educación primaria del país. A su vez, estos libros son una compilación de textos literarios de diferentes autores que corresponden con la tradición cultural universal y del país a los que están expuestos la mayoría de los niños mexicanos.

El Corpus se compuso por un conjunto de 604 textos en su mayoría literarios (526) como se puede apreciar en la Tabla 1, divididos en 6 libros correspondientes a libros de texto de lectura para cada grado en sus ediciones 1993, 1997, 2011 y 2012 y dos antologías también para cada grado en su edición de 2011.

Tabla 1. Géneros Textuales que Componen el C-TEN

Total de la muestra	100%	Porcentaje por género	100%
Literarios	79.97	Cuentos	53.15
		Canciones, rimas, adivinanzas, juego de palabras, coplas	11.92
		Poemas	8.11
		Leyendas	3.31
		Fábulas	2.81





		Cuentos adaptados a obras de teatro	0.66
Informativos	19.7	Monografías	19.7
Otros	0.33	Crónica	0.16
		Instructivo	0.16

El Corpus de Textos Escritos para Niños en su conjunto lo constituyeron 141,230 palabras totales (tokens) de los que resultaron 14,572 palabras diferentes o gráficas (types). Las 14,572 palabras gráficas (types) se clasificaron de acuerdo con su naturaleza en léxicas (o de contenido) y de función (o gramaticales)¹. La categoría gramatical con mayor número de palabras diferentes o tipos fueron los verbos con 6,067 types con una ocurrencia de 27,616 tokens.

En la identificación del vocabulario fundamental obtuvimos del proceso de lematización 1260 cabezas léxicas de las cuales resultaron 185 lemas con una ocurrencia del 15,071.

b) Resultados de la identificación de la clasificación semántica

En cuanto a la clasificación semántica del vocabulario fundamental encontramos que el nivel máximo de concentración fueron los procesos material y mental.

Como se puede observar en la Tabla 2, de las seis macro-clases el mayor porcentaje de ocurrencias (64.42%) se concentra en los procesos material y mental. A este nivel lo denominamos superior para desarrollar a continuación la caracterización de las clases y subclases. El proceso material representa el 42.09% siendo el más frecuente y el mental, que es el que sigue, abarca el 23.33% del total. Cabe aclarar que la macro-clase material: es la categoría gramatical de los procesos que expresan las experiencias del mundo exterior. Por su parte, el proceso mental: es la categoría gramatical de los procesos que expresan las experiencias del mundo interior. En oposición de los procesos menos frecuentes que tienen que ver con aprender a generalizar – relacionar un fragmento de la experiencia con otro: categorizar. En gramática se reconocen los procesos que expresan identificación y clasificación.

¹ Para conocer el procedimiento, herramientas y listado de las palabras del corpus según la categoría gramatical, consultar García (2014). En este estudio únicamente presentamos el inventario de verbos.





Tabla 2. Tipo de Proceso, Número de Lemas, Frecuencia y Porcentaje Acumulado

Etiqueta (tipo de proceso)	N lemas	Frecuencia	%	% acumulado
Material	97	6044	42.09	42.09
Mental	33	3208	23.33	64.42
Verbal	15	2364	16.45	80.87
Existencial	17	1246	7.43	89.51
Relacional	17	1067	7.43	96.94
Modulación	6	440	3.06	100.00

A continuación describimos la agrupación superior.

NIVEL SUPERIOR DE CONCENTRACIÓN: MACRO-CLASES MATERIAL Y MENTAL





La diferencia en cuanto al número de lemas y ocurrencias entre las macro-clases material y mental es amplia. La macro-clase material (97 lemas) supera por 64 lemas a la macro-clase mental (33 lemas). También en el número de ocurrencias la macro-clase material superó casi al doble a la macro-clase mental. Presentaron 6,044 y 3,208 ocurrencias respectivamente. Por lo que podemos concluir que la macro-clase material concentra el mayor porcentaje de acumulación tanto de lemas (52.42%) como de ocurrencias (42.09%) de los verbos más frecuentes del Corpus de Textos Escritos para Niños.

MACRO-CLASE MATERIAL

En el marco conceptual amplio de la macro-clase de verbos materiales, que resultó ser el de mayor concentración del corpus, identificamos las clases de espacio, otros hechos, cambio y comportamiento. En el vocabulario fundamental del Corpus, los verbos de espacio son los que presentan, a su vez, el mayor número tanto de lemas (48) como de ocurrencias (3006). Este elevado número de ocurrencias representa el 24.91% del vocabulario fundamental. Seguido de los verbos caracterizados como de espacio, en orden de mayor a menor representatividad en la macro-clase material, encontramos los verbos de comportamiento (13 lemas y 952 ocurrencias), de otros hechos (20 lemas y 819 ocurrencias) y finalmente los de cambio (16 lemas y 690 ocurrencias). En la siguiente tabla presentamos algunos ejemplos de verbos que pertenecen a cada una de las clases que se agrupan en la macro-clase material.

Tabla 3. Clases de la Macro-clase Material: No. Lemas, Frecuencia y Porcentaje

Clase	N lemas	Frecuencia	% de la macro-clase	Verbos (ejemplos)
Espacio	49	3580	59.23	<i>llegar, meter, levantar, entrar, sentar</i>
Comportamiento	13	952	15.75	<i>casar, dormir, llorar, despertar</i>
Otros hechos	20	819	13.60	<i>jugar, picar, agarrar, usar, sonar</i>
Cambio	16	690	11.42	<i>abrir, cortar, formar, tejer, lavar</i>





De los verbos de la macro-clase material clase espacio, que son los de mayor ocurrencia, se agrupan a su vez en subclases que son: desplazamiento, localización, postura-posición y unión. La subclase en la que se concentraron el mayor número de lemas fueron los verbos de desplazamiento en la que encontramos 34 lemas con 3006 ocurrencias que representan el 49.73% de la macro-clase y el 20.92% del vocabulario fundamental. Algunos ejemplos fueron: *llegar, salir, llevar, venir y volar*.

Las subclases de la clase denominada de comportamiento son: relaciones, fisiología y de intersección con otra clase. Los lemas que se presentaron con mayores ocurrencias de las subclases son los que comparten elementos del marco conceptual con respecto a la fisiología, encontramos por ejemplo: *comer, dormir, llorar, reír y parpadear*. Todavía es posible encontrar elementos comunes en estos lemas, por ejemplo, *dormir, llorar, reír y parpadear* son prototípicos de esta subclase y *comer* tiene que ver con ingestión como *beber y alimentar*.

La clase siguiente de acuerdo al criterio de mayor a menor frecuencia es la que agrupa los otros hechos. Las subclases de los otros hechos son: actividad, contacto, control, uso, emisión y competición. En la subclase que identificamos más ocurrencias fue en la de actividad que agrupa a 4 lemas con 312 ocurrencias. Los lemas de esta subclase son: *jugar, trabajar, preparar y bailar*.

La última clase de la macro-clase material que es la que presentó menor número de ocurrencias es la de cambio que representa el 7.61% con respecto al grupo de referencia (macro-clase). Las subclases son: de modificación y de creación. La subclase con mayor número de lemas (11) y ocurrencias (511) fue la de modificación. Algunos ejemplos de lemas que encontramos en esta subclase son: *abrir, cortar, crecer, secar y cambiar*.

Para concluir con la parte de resultados, a través de la Figura 1 que tomamos de Halliday y Matthiessen (2004) tratamos de representar las frecuencias de las clases semánticas de los verbos del vocabulario fundamental en la gramática de la experiencia. Es decir, el máximo nivel de concentración lo encontramos en los procesos (material 42% y mental 23%) que están relacionados con los mundos físico y de la conciencia respectivamente. Que son los de adquisición más temprana en la oralidad. Le sigue el proceso verbal 16% que se ubica en la





frontera entre el proceso material (pensar) y el de simbolizar incluido en los proceso relacional y de modulación. A continuación el proceso existencial con el 8.6% y el relacional y modular que en conjunto representaron el 10% que están relacionados con el mundo de las relaciones abstractas.

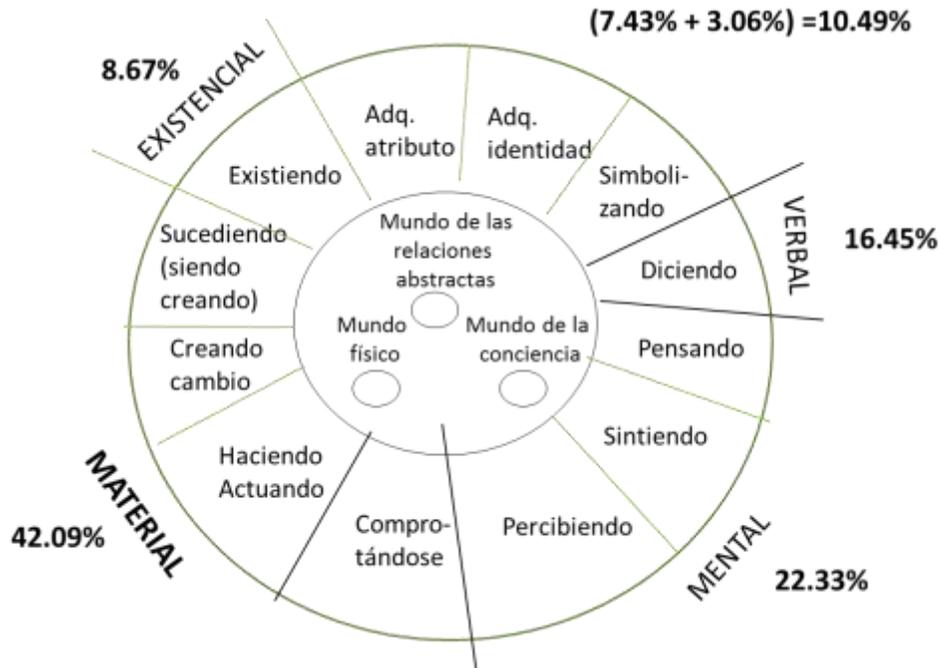


Figura 1. Representación Semiótica de la Gramática de la Experiencia del Vocabulario Fundamental

Adaptación del esquema Gramática de la experiencia: tipo de proceso de Halliday, y Matthiessen (2004) para la presentación de los resultados de la clasificación semántica..

CONSIDERACIONES FINALES

Como se mostró los resultados, los procesos relacionados con la gramática de la experiencia que fueron los más frecuentes estuvieron asociados al mundo físico y de la conciencia. Las categorías de uso restringido fueron las que dan cuenta del mundo de las relaciones abstractas. Parecen lógicos estos resultados si consideramos que los textos de donde los obtuvimos están escritos para niños en momentos iniciales de alfabetización inicial. Lo que resulta relevante recalcar es que no son textos programados sino una compilación de textos escritos por varios autores que representan la tradición cultural de la humanidad y de México.





Adicionalmente a que este análisis lograra facilitar datos que pueden resultar eventualmente útiles para estudios cuyo interés sean verbos escritos en estudios psicolingüísticos, evaluación del lenguaje y la toma de decisiones didácticas para favorecer la adquisición de la lengua escrita. Nos parece que en sí mismo, los procedimientos utilizados para realizar el estudio pudiera dar pistas a los docentes o formadores de docentes de estos grados escolares (1º y 2º de primaria) de posibles formas de análisis lingüístico en el que se pudieran involucrar los docentes para que a la vez que problematizan la complejidad del español también surjan ideas para la planeación de situaciones didácticas que realmente planteen conflictos cognitivos que permitan a los niños construir o re-construir la lengua escrita.

La última consideración tiene que ver con los futuros estudios que se pudieran desarrollar a partir de este corpus de verbos. De acuerdo al objetivo del presente estudio: Proporcionar un análisis de los verbos contenidos en el Corpus de Textos Escritos para Niños con la intención de facilitar datos que puedan resultar eventualmente útiles para estudios psicolingüísticos. Los distintos inventarios con criterios específicos elaborados podrían facilitar recursos para estudios futuros: A partir del propio estudio, bajo la consideración de que este estudio fue de carácter exploratorio y general por lo que se abren varias líneas de análisis, por ejemplo: se podría realizar un análisis en contexto de los verbos con potencial metafórico para identificar las locuciones verbales a los que están expuestos los niños de estos grados escolares. Desde la psicolingüística: estudios cognitivos relacionados con el reconocimiento de palabra, lectura de palabras en contexto, comprensión y fluidez lectora. Desde la lexicografía continuar con descripciones útiles de los contextos del corpus, por ejemplo, descripciones sobre las perífrasis verbales como un fenómeno funcional y semántico interesante. Desde la psicología educativa, los distintos listados, pueden resultar útiles para la toma de decisiones relacionadas a la adquisición de la lengua escrita.





BIBLIOGRAFÍA Y REFERENCIAS

- Dixon, R. (1991). *A New Approach to English Grammar, on Semantic Principles*. Oxford: Oxford University Press.
- Faber, P. y Mairal, R. (1999). *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.
- Fellbaum, C. (1998). *A Semantic Network of English Verbs*. In C. F. (ed), *WordNet. An Electronic Lexical Database* (pp. 69-104). Cambridge (Massachusetts): The MIT Press.
- Fillmore, C. J. (16/ 3, 2003). *Background to FrameNet*. *International Journal of Lexicography*, 235-250.
- García Aldeco, A. (2013). *Corpus de sustantivos más frecuentes en textos escritos para niños mexicanos, en momentos iniciales de alfabetización tesis inédita*. (Tesis inédita de maestría). Universidad Autónoma de Querétaro, Querétaro, Qro.
- García-Miguel, J. M., Costas, L. y Martínez, S. (2003). *Diateisis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE*. Leipzig.
- Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.
- Halliday, M.A.K. (2004b). *Lexicology*. In M. T. Halliday, *Lexicography and Corpus Linguistics* (pp. 1-22). London-New York: Continuum.
- Halliday, M.A.K. y Matthiesen, C. M. (2004). *An Introduction to Functional Grammar*. London-New York: Hodder Arnold.
- Halliday, M. y Matthiessen, M. (2006). *Constructing experience through meaning: a language-based approach to cognition*. London-New York: Continuum.
- Lara, L.F., Ham, C y García, M (1979). *Investigaciones lingüísticas en lexicografía*. México: El Colegio de México.
- Levin, B. (1993). *English Verb Classes and Verbal Alternations. A Preliminary Investigation*. Chicago: University of Chicago Press.





Sinclair, J. (2005). Corpus and Text - Basic Principles. En M. Wynne (ed.) Developing Linguistic Corpora: a Guide to Good Practice (pp. 1-16). Consultado en <http://ahds.ac.uk/linguistic-corpora/>

Vázquez, G., Ana, F. y Martí, M.A. (2000). Clasificación verbal. Alternancia de diátesis. Lleida: Universitat de Lleida.

ADESSE consultada en <http://webs.uvigo.es/adesse/>

ARTHUS consultada en <http://www.bds.usc.es/>

BDS consultada en <http://www.bds.usc.es/corpus.html>

FrameNet consultada en <http://sfn.uab.es:8080/SFN>

WordNet consultada en <http://wordnet.princeton.edu/>

