

# **ANALISIS MULTIDIMENSIONAL DE TEXTOS ESCRITOS**

EN ESPAÑOL DE VENEZUELA

Elizabeth ArcayHands y Ligia Cossé

Universidad de Carabobo. Valencia, Venezuela

C.E.: earcay@thor.uc.ve / cperez@thor.uc.ve

Los objetivos de la línea de investigación ANÁLISIS MULTIDIMENSIONAL DE TEXTOS ACADEMICOS ESCRITOS, una de las líneas de investigación adscritas al proyecto para la creación del Centro de Investigaciones en Lingüística Pura y Aplicada de la UC, se centran en dos grandes aspectos. El primero, en el análisis de textos escritos en español de Venezuela por sujetos monolingües y bilingües. El segundo, en la detección de semejanzas y diferencias que informen sobre procesos de transferencia del inglés como L2 al español de Venezuela como L1, en lo que a lenguaje escrito se refiere. A continuación se presenta un descripción de la línea en cuanto a su origen, desarrollo y resultados.

## **Introducción**

Las últimas tendencias en análisis del discurso escrito (Moore y Pollack, 1992; Hatch, 1993; Biber, 1993; Biber, Conrad y Reppen, 1994) plantean que la interacción de aspectos lingüísticos - puros y textuales -, psicológicos, sociales y culturales en los textos escritos es empíricamente identificable. Así, una caracterización global de textos debe, necesariamente, provenir de tantas dimensiones como sea posible para obtener una descripción más cónsona con la naturaleza textual. Por otro lado, diferencias reportadas en estudios de la producción escrita de estudiantes monolingües y bilingües (español-inglés) en diferentes dimensiones textuales (Montaño-Harmon, 1991, Reppen y Grabe, 1993) apuntan hacia la necesidad de una exploración más profunda de los resultados de procesos de transferencia no sólo de una L1 a una L2 sino también de una L2 a una L1 (Arcay, 1996). Asimismo, apuntan hacia la necesidad de analizar textos producidos no sólo por estudiantes sino por sujetos profesionales y dirigidos a un público mayor.

## **La multidimensionalidad**

En la literatura del área existe un abanico bastante amplio en cuanto a los tipos de análisis susceptibles de ser seleccionados para explorar las cinco dimensiones textuales anteriormente mencionadas. De estas posibilidades se hizo una selección de cinco tipos de análisis, los cuales se corresponden con cada una de las dimensiones. Para el estudio de los aspectos lingüísticos se utilizaron las unidades T (Hunt, 1965), la clasificación de oraciones ortográficas de E. Arcay (1996) y la presencia de verbos principales en elipsis. Para el estudio de los aspectos lingüísticos textuales se utilizó la caracterización de los patrones léxicos de M. Hoey (1991). Para el estudio de los aspectos cognitivos se utilizó como unidad de análisis la identificación de las proposiciones del texto de acuerdo a los bloques discursivos de Pitkin (1969). Para el estudio de los aspectos culturales, se desarrolló una unidad de análisis, basada en los planteamientos de R. Kaplan (1983, 1984, 1987), que caracteriza la organización retórica, a través de la detección de unidades discursivas (ud) y de unidades retóricas (ur). Finalmente, para el estudio del aspecto social se utilizó la interacción productor-receptor, a través de la detección de ítemes de metadiscurso interpersonal (Vande-Kopple, 1991; Crismore and Farnsworth, 1990).

## **El corpus**

Hasta el presente, el corpus está constituido por tres sub-corpora. El primero está formado por diecisiete (17) textos, tipo ensayo académico, escritos en español de Venezuela, producidos por estudiantes monolingües y bilingües (español e inglés) quienes cursaban el último semestre de las Licenciaturas en Educación, Menciones Español como Lengua Materna e Inglés como Lengua Extranjera. El segundo está constituido por cinco textos tipo ensayo académico, susceptibles de ser considerados modelos de su género (la selección de textos susceptibles de ser considerados modelos de su género constituye una de las etapas principales para la caracterización de un género textual (Bhatia, 1993)). De los cinco textos, tres están escritos en español de Venezuela, uno por un sujeto bilingüe (español de Venezuela e inglés) y dos por uno monolingüe, los dos restantes están escritos en inglés. El tercer sub-corpus, actualmente bajo análisis, está constituido por artículos de opinión escritos por autores monolingües del español de Venezuela,

publicados en la prensa regional, durante el año 1996. La selección de ensayos académicos y de artículos de opinión responde al hecho de que, a nuestro entender, los primeros constituyen una de las formas más frecuentes de producción escrita en contextos académicos. Los segundos por ser textos de extensión manejable y de fácil acceso para su análisis.

En la creación del corpus de ensayos académicos producidos por autores reconocidos, hablantes de español como L1, surgieron dificultades que son importantes de mencionar. Durante la etapa de recolección de textos para su posterior selección fue sumamente difícil encontrar textos cuya extensión oscilara entre 30 y 60 oraciones, extensión ideal para la aplicación de las unidades por dos razones. La primera, porque este número de oraciones es el mínimo necesario para la comprobación de las propiedades del análisis de patrones léxicos. Y la segunda, porque permite el análisis manual de las unidades que todavía no se han podido automatizar. Por otra parte, la mayoría de los textos encontrados o bien poseían la estructura organizativa de un artículo de revista especializada, o constituían el capítulo de un libro -imposible de ser analizado independientemente- o eran artículos de prensa. que no cumplían los requisitos necesarios para ser considerados textos académicos. Es decir, una composición escrita, independiente de otros textos y realizada por estudiantes universitarios o por un autor con el propósito de ser publicada; en ella se expresa la opinión del autor en relación al tópico tratado y está dirigida a una audiencia ubicada dentro de un entorno universitario (Arcay y Cossé, 1997).

## **Las Unidades de Análisis**

En relación a la capacidad descriptiva de cada una de las cinco unidades de análisis por separado, podemos decir, en primer lugar, que la clasificación de las oraciones ortográficas (dimensión lingüística) se constituye en una herramienta poderosa que permite una descripción detallada de las múltiples combinaciones de cláusulas hipotácticas y paratácticas que se dan en textos reales. Esta descripción detallada consiste en la detección de cuatro tipos de oraciones adicionales a los tres tipos descritos en la gramática española (GRAE, 1985; Bello, 1984; Seco, 1972; Marin, 1972). Además de las oraciones simples,

coordinadas y subordinadas, esta clasificación describe cuatro combinaciones posibles de oraciones subordinadas (subordinada simple, subordinada en coordinación, subordinada en subordinación y subordinada mixta) y una para el uso simultáneo de la coordinación y la subordinación.

Las siete combinaciones propuestas en la clasificación permiten no sólo una descripción más precisa sino también abrir la posibilidad de dar explicaciones a ciertas pre-concepciones en cuanto a la longitud y al tipo de oraciones ortográficas favorecidas en cada idioma. A priori, es común el considerar que en español las oraciones son más largas (Santana-Seda,1975) y más complejas que en inglés. En el corpus correspondiente a los textos modelos (noasi en los textos escritos por estudiantes) se detectó que la diferencia entre la longitud promedio en ambos idiomas no fue significativa y que el tipo de oración más frecuente fue el de las oraciones simples, a pesar de las variaciones en frecuencia. Sin embargo, se detectó que en español el mecanismo de combinación de cláusulas preferido fue el de la subordinación. Y en inglés, por el contrario, fue el de la coordinación. Por otra parte, en los textos producidos por el sujeto monolingüe se detectó una marcada preferencia por sólo dos tipos de oraciones, en cambio en los del sujeto bilingüe hubo una distribución más heterogénea. Es decir, se detectaron oraciones de todos los tipos.

Entonces, si no fue posible detectar diferencias reales en la longitud promedio de las oraciones, sino en el favorecimiento de combinaciones hipotácticas de cláusulas, pensamos que la concepción de una mayor longitud oracional en español vendría a estar dada no por el número de palabras que las constituyen, sino porque las cadenas de subordinación inducen a pensar que ellas están compuestas por un mayor número de palabras. Por otro lado, el uso más frecuente de sólo dos tipos de oraciones podría también contribuir a la percepción de mayor longitud y/o complejidad. Pensamos que los hallazgos detectados indican la necesidad de una exploración más profunda de este aspecto con otros corpora.

En lo concerniente al análisis de patrones léxicos (dimensión lingüística textual), éste permite acumular evidencia para la caracterización de los tipos de repetición léxica en español -hecho que es de suma importancia para la descripción de nuestra lengua por cuanto el uso de la repetición

ha sido hasta ahora cuestionado-. Sin embargo, el hecho de que tres características de los textos de el autor monolingüe, ocasionaron ciertos problemas a la hora de la comprobación de algunas de las propiedades del análisis, merece ser mencionado. Estas características fueron el uso de la elipsis en un porcentaje muy elevado, el uso de figuras literarias en textos que se pensaban que eran puramente académicos y la presencia de un segmento narrativo bastante largo en uno de los textos. Esto se tradujo en que, por un lado, se produjera un número muy bajo de repeticiones por conexión y de conexiones, y por el otro, una aparente incoherencia en momento de la elaboración de los resúmenes con las oraciones pertenecientes al segmento narrativo del texto.

En vista de que es la primera vez que ocurre este tipo de problema con el análisis, pensamos que lo anterior no constituye evidencia en contra de sus bondades sino que las dificultades tienen que ver con las características mencionadas, las cuales no son armónicas con los mecanismos que lo gobiernan. Tal y como lo señala su autor (Hoey,1991) el análisis no fue diseñado para textos narrativos; por otra parte, la creación de conexiones a través de repeticiones por elipsis ocupa el lugar más bajo en la jerarquía de repeticiones; es decir, es la que va a producir el menor número de conexiones. Asimismo, el uso de metáforas no permite la creación de conexiones por cuanto es prácticamente imposible detectar una repetición de los ítemes léxicos que las conforman fuera del contexto inmediato donde aparecen. Es sumamente importante entonces analizar otros textos escritos por hablantes monolingües del español de Venezuela para comprobar si la presencia de la elipsis y de las figuras literarias en textos académicos es una característica común del español o si simplemente es idiosincrásica del autor monolingüe analizado.

El análisis de los bloques discursivos (dimensión cognitiva) permite la detección de la cantidad y complejidad de las ideas utilizadas para sustentar las proposiciones de un texto. A su vez ello permite que se le utilice como herramienta para comprobar la densidad y profundidad de la sustentación y el soporte de ideas.

Por su parte, aunque el análisis de la organización retórica (dimensión cultural) permitió el acceso a información importante, no pudo ser capaz de detectar diferencias que de una manera u otra se presumía como ciertas. Pensamos que una de las razones que ocasionaron este

problema es la ausencia de indicadores explícitos de las unidades, ya que la mayoría de las veces los autores no presentan explícitamente indicadores de las relaciones lógicas que se crean entre las proposiciones de sus textos, bien sea por una razón o por otra. Como consecuencia, al no ser posible detectar ítemes explícitos y tener que hacer uso de la interpretación, el análisis toma un carácter subjetivo que no posee el resto de las unidades aplicadas. Es importante entonces tratar, en lo posible, de afinar el procedimiento para la detección combinaciones de elementos que representen a dichas relaciones y, por ende, permitan caracterizar a las unidades de forma más objetiva. Finalmente, el análisis del metadiscurso (dimensión social) permite detectar características importantes en cuanto al tipo de ítemes favorecidos.

En general, la capacidad del análisis multidimensional para suministrar información de naturaleza diferente de forma simultánea se evidencia al permitir una descripción mucho más rica, acorde con la complejidad textual del lenguaje escrito. Pensamos que es sumamente importante continuar creando corpora de textos escritos en español de Venezuela de mayores dimensiones y aumentar la automaticidad de la aplicación de las unidades para poder ampliar el alcance de los resultados obtenidos.

### **Direcciones Futuras:**

El análisis multidimensional de textos nos ha permitido dar un paso más en la descripción de textos de lenguaje natural escritos en español de Venezuela, con la detección de diferencias y semejanzas en los textos del corpus analizado. El análisis permitió también detectar las debilidades y fortalezas de cada una de las unidades seleccionadas para el estudio de las cinco dimensiones textuales y con ellos posibles correctivos para futuros análisis. Los pasos que nos proponemos seguir en esta línea de investigación son:

1. Crear un corpus del idioma español de Venezuela LN de dimensiones mayores, tal y como lo sugieren los resultados de investigaciones realizadas en el idioma inglés (Baker, Francis, y Tognini-Bonelli, 1993).

2. Desarrollar herramientas de análisis computarizadas, o adaptar aquellas existentes, que permitan acelerar y aumentar el alcance de las unidades de análisis aquí utilizadas.

3. Considerar el bilingüismo como una variable de control a la hora de seleccionar textos para la creación de los sub-corpora.

### **Referencias:**

Arcay, E. (1996). Análisis multidimensional de textos académicos escritos en español de Venezuela. Trabajo de grado no publicado. Universidad Simón Bolívar.

Arcay, E. y Cossé L. (1997). Ensayos académicos: Un análisis multidimensional de textos escritos en español de Venezuela e inglés. Ms. Universidad de Carabobo, Venezuela.

Baker, M., Francis, G. y Tognini-Bonelli, E. (Eds.) (1993). Text and Technology. In honour of John Sinclair. Amsterdam: John Benjamins Publishing Company.

Bhatia, V. J. (1993). Analysing genre: Language use in professional settings. London: Longman.

Biber, D. (1993). The multidimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computer and the Humanities*, 26, 331-345.

Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15(2), 169-189.

Crismore, A., & Farnsworth, R. (1990). Metadiscourse in popular and professional science discourse. En W. Nash (Ed.), *Written Communication Annual: Vol. 3. The Writing Scholar: Studies in Academic Discourse* (pp. 118-136). Newbury Park, California : Sage.

Hatch, E. (1992). *Discourse and language education*. Cambridge: Cambridge University Press.

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hunt, K. (1965). *Grammatical structures written at three grade levels*. (Research Report N° 3). Urbana, Illinois: National Council of Teachers of English.

Kaplan, R. B. (1983). *Contrastive rhetorics: Some implications for the writing process*. En A. Freedman, I. Pringle, & J. Yalden (Eds.), *Learning to write: First language/Second language* (pp. 139-161). Selected papers from the 1979 CTE Conference, Ottawa, Canada. New York: Longman.

Kaplan, R. B. (1984). *Cultural thought patterns in inter-cultural education*. En S. McKay (Ed.), *Composing in a Second Language* (pp. 43-62). Cambridge, Massachusetts: Newbury House.

Kaplan, R. B. (1987). *Cultural Thought Patterns Revisited*. En U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 9-22). Reading, Massachusetts: Addison-Wesley.

Montaño-Harmon, M.R. (1991). *Discourse features of written Mexican Spanish: Current research in contrastive rhetoric and its implications*. *Hispania*, 74(2), 417-425.

Moore, J.D. y Pollack, M.E. (1992). *A problem for RST: The need for multilevel discourse analysis*. *Association for Computational Linguistics*. 18 (4) 537-544.

Reppen, R., & Grabe, W. (1993). *Spanish transfer effects in the English writing of elementary school students*. *Lenguas Modernas, Universidad de Chile*, 20, 113-128.

Santana-Seda, Sr. O. (1975). *A contrastive study in rhetoric: An analysis of the organization of English and Spanish paragraphs written by native speakers of each language*. Unpublished doctoral dissertation. New York University.

VandeKopple, W.J. (1991). *Themes, thematic progressions, and some implications for understanding discourse*. *Written Communication*, 8(3),

311-347.