

# AVOCES: Sistema para el etiquetado de discursos orales

Luis Carlos Toro Tamayo\*

## Resumen

La comunicación oral es natural al ser humano y constitutiva de la persona como miembro de una especie. Se produce en y con el cuerpo, aprovechando órganos del sistema respiratorio y de la deglución. También los movimientos de los ojos, diferentes expresiones faciales y otros movimientos corporales forman parte importante de la oralidad, así como las “vocalizaciones” (sonidos bucales aunque no lingüísticos) y otros “ruidos”. Sin embargo, una conferencia, un discurso inaugural o un juicio, por ejemplo, aunque manifestaciones orales todas, requieren un alto grado de preparación e incluso, muchas veces, el uso de la escritura (un guión, unas notas, un acta, un expediente, etc.). Llegar a dominar esas formas de hablar no es sencillo y por eso el desarrollo de la competencia comunicativa oral es también parte de la educación lingüística. Según lo anterior, este texto tiene como objetivo presentar las decisiones tomadas en relación con la definición de categorías útiles para el análisis de los discursos orales, específicamente las relacionadas con el etiquetado y la transcripción. Dichas orientaciones están acompañadas de premisas teóricas que las sustentan y el diseño de un algoritmo que organice los procesos de sistematización y almacenamiento de este tipo de fuentes.

**Palabras clave:** Análisis de discursos orales. Tecnología Lingüística. Etiquetado y transcripción.

## 1 Introducción

La investigación Sistema para el Etiquetado de Discursos Orales, busca definir algunas categorías útiles para el análisis de este tipo de fuentes, basándose en premisas teóricas y metodológicas ampliamente sustentadas. Dicho proyecto tuvo previsto el diseño de un algoritmo denominado AVOCES, que agiliza los procesos de sistematización y almacenamiento documental.<sup>1</sup>

---

\* Historiador, Magíster en Lingüística (Univ. de Antioquia – Colombia). Investigador del Departamento de Ciencias Sociales y Humanas (Univ. de Medellín – Colombia). Doctorado (en curso) en Estudios Latinoamericanos (Univ. de Chile – Programa Becas Latinoamericanas para Estudios de Doctorado en Chile 2008 – Conicyt). E-mail: karlostoro@hotmail.com.

<sup>1</sup> El proyecto *Sistema para el Etiquetado de Discursos Orales* surge en 2005 por iniciativa de los grupos de investigación de la Universidad de Medellín *Phrónesis*, del Departamento de Ciencias Sociales y Humanas, y *Arkadius*, de la Facultad de Ingeniería. Actualmente hacen parte del proyecto el Historiador y Magíster en Lingüística Luis Carlos Toro Tamayo (Investigador Principal), las Magíster en Ingeniería de Sistemas Diana María Montoya Quintero e

Se trata de un estudio multidisciplinar que intenta, mediante enfoques como el Análisis del Discurso y la Tecnología Lingüística, fortalecer las técnicas de sistematización de los discursos orales y su aplicación en campos como la historia, la antropología, la sociología, la lingüística y el derecho, entre otras.

Creemos que es menester de los investigadores, y de los centros de investigación, idear métodos alternativos y más confiables de extracción y análisis de la información, que nos permita estudiar los discursos y a la vez construir documentos que contribuyan con la preservación de la tradición oral de nuestros antepasados.

## 2 Presupuestos y autores de referencia

Los discursos orales tienen tantas particularidades, diferentes a las del discurso escrito, que son muchas las ocasiones en las que el alfabeto no nos alcanza para interpretarlas. Las diferencias se observan, no sólo en lo referente a la variación fonológica, sino en la segmentación de los signos, las interrupciones, las acotaciones, los silencios, etc. Es aquí donde se vuelve indispensable la definición de categorías de análisis, con atributos homogéneos para la mayoría de investigadores, que permita una correcta interpretación de los discursos orales y que tenga la virtud de conservar el material impreso y digital bajo un método confiable.

Dicho asunto ha generado gran polémica entre quienes piensan que la puntuación traiciona al texto oral, y que por consiguiente, no sólo afecta su sentido, sino que puede conducir a un análisis equivocado (BLANCHE-BENVENISTE, 1998). Por tal motivo, el objetivo de la presente investigación consistió en desarrollar un sistema que permitiera hacer análisis más acertados de discursos orales, a partir de la comparación del original y la interpretación realizada por el investigador. Esto es, un algoritmo especializado en la extracción de los datos del sistema, que permita analizar los discursos orales de manera ágil y confiable.

Dentro de las pautas generales que tuvimos en cuenta para la transliteración de las conversaciones y sobre los aspectos fonéticos del discurso conversacional, están los aportes realizados por Briz A. y J. R. Gómez Molina (1992), Sacks, Schegloff y Jefferson (1974), Schenkein (1978), López Morales (1994) y Levinson (1983).

---

Idanis Beatriz Díaz Bolaño (Co-Investigadoras), y las estudiantes Catalina Peláez Escobar y Laura María Pareja Georges (auxiliares), de la Facultad de Derecho y el Programa de Ingeniería de Sistemas, respectivamente. La fuente de financiación del proyecto proviene de fondos destinados por la Vicerrectoría de Investigación de la Universidad de Medellín.

Lo que respecta al análisis del discurso retomamos los planteamientos de Helena Calsamiglia y Amparo Tusón (2002), Jesús Tusón (1997), Violette Morin (1969, 1974), Teun A. Van Dijk (1990, 1991, 1995, 1997, 1998, 2000, 2003), J. Renkema (2000), F. Van Eemeren y Otros (2002), Anthony Weston (1994), Ruth Wodak (2003), Michael Meyer (2003), Siegfried Jäger (2003), Norman Fairclough (2003) y Ron Scollon (2003), entre otros.

Y es que para comprender los discursos orales debemos tener en cuenta el conjunto de personas e informaciones que participan en la conversación, en el que aparecen diversos grados de información, dependiendo del lugar, las circunstancias sociales, políticas, económicas, etc., haciendo una clara distinción entre el texto y su contexto, tal y como lo plantea el método de Análisis de la Información Documental (MORIN, 1969, 1974), y el enfoque de Análisis Crítico del Discurso, ACD, (VAN DIJK, 1990, 1991, 1995, 1997, 1998, 2000, 2003).

De este modo, análisis del discurso, recuperación de la memoria oral, ideología, cultura y sociedad, hacen parte de un terreno común de la investigación, que merece ser analizado bajo parámetros sistemáticos por sociohumanistas e ingenieros, a fin de garantizar la validez de las afirmaciones realizadas en las investigaciones y comprobar la legitimidad de los mismos con elementos de juicio, respaldados por la observación empírica.

### **3 Principios para la conformación de corpus orales**

La configuración de corpus orales no consiste en la simple acumulación de materiales, sino en la reflexión sobre los aspectos inherentes a la representación de los datos. Dicho estudio ha sido definido con el concepto de Etiquetado de los discursos orales, e implica un arduo proceso, máxime si tenemos en cuenta lo limitado de los tiempos en la investigación y que, entre transliterar, etiquetar y revisar el material, podemos ocupar diez veces el tiempo de las grabaciones.

Según Amparo Tusón Valls (2003), solo hasta ahora existe un creciente interés por el estudio de los discursos orales. Dichas reflexiones se han convertido en centro de atención de diversas perspectivas científicas, entre las que se destacan la lingüística, la comunicación y las ciencias sociales y humanas. Ya en lo que respecta a la fusión de disciplinas como la lingüística y la ingeniería de sistemas, podemos apreciar una evolución en el terreno de las tecnologías lingüísticas mediante el programa Euromap. Dicho estudio nos muestra la cantidad de información sobre el estado de investigación y desarrollo de las tecnologías lingüísticas en cada país miembro de la Unión Europea.

Otras disciplinas y subdisciplinas como la antropología, la sociología, la ciencia cognitiva, la filosofía, la sociolingüística, el análisis del discurso, la lingüística textual o la pragmática, comparten un interés por la interacción verbal, en aras de la definición de metodologías de procesamiento del lenguaje natural y la concepción de teorías específicas sobre el tema.

Y es que por lo general los registros orales tienen que ser escuchados y analizados desde una perspectiva interdisciplinar, atendiendo a puntos de vista como: el control de las estructuras de los temas y de los contenidos, el control de la estructura sintáctica, el control de la superestructura, el nivel léxico, el control de las formas retóricas, la distribución de los turnos, la clasificación de los participantes, los juegos del implícito y el explícito en la interacción comunicativa, la ideología, el contexto, los aspectos prosódicos y semióticos, entre otros.

Según Régis Debray (2001, p. 15), hablar no consiste sólo en transmitir un mensaje preparado con antelación y mediante una técnica adecuada, sino condensar la memoria colectiva de un grupo histórico que perpetúa a través de los años una "personalidad de base", y que hace parte de el momento de un proceso que será más largo y el fragmento de un conjunto más vasto.

Así, para sacar adelante el sistema AVOCES, hicimos énfasis en la transcripción ortográfica para pasar luego a un nivel de representación fonética mucho más profundo, tanto en el aspecto segmental como en el suprasegmental<sup>2</sup>. También, planteamos la cuestión del etiquetado de los discursos que, asociado a otras operaciones como la alineación temporal, permite una utilización posterior del corpus en diversas áreas del conocimiento.

## **4 Pasos que debemos cumplir en el etiquetado de discursos orales**

Si bien existen diferentes propuestas de categorías para el etiquetado de los corpus orales, entre las que se destacan autores como Sacks et al., (1974), Gumperz y Berenz (1990), Du Bois (1991), Payrató (1995), Llisterri (1999), Briz A. y el Grupo Val.Es.Co (2000), retomamos las orientaciones hechas por Amparo Tusón Valls (2003), quien propuso un método de transcripción para que los materiales orales sean objetos manejables, susceptibles de un tratamiento analítico.

---

<sup>2</sup> Se entiende por suprasegmental todo elemento que afecta a más de un fonema y que no puede segmentarse en unidades menores, como el acento, la entonación o el ritmo.

Según Payrató (1995, p. 45), la transcripción es un procedimiento de adaptación de una producción lingüística oral a una forma gráfica escrita. De este procedimiento hacen parte la transliteración o transcripción ortográfica y la transcripción fonética, que se lleva a cabo mediante signos diseñados para tal propósito. No obstante, dicho proceso de traducción de un lenguaje oral a uno escrito requiere un proceso de reflexión y análisis, en el que se consideran aspectos como el contexto, cuyo significado social depende de la actividad humana y el contexto comunicativo, entre otras. La competencia comunicativa es pues un campo de estudio que ha venido avanzando y que apunta más al desarrollo de los aspectos culturales y situacionales del uso del lenguaje, que a las estructuras gramaticales y al vocabulario de las personas.

Entre las categorías de análisis que utilizamos en nuestro Sistema para el Etiquetado de los Discursos Orales, tuvimos en cuenta aspectos concernientes a los códigos lingüísticos, paralingüísticos y extralingüísticos (NIÑO ROJAS, 2003, p. 57-59), justificados en los métodos técnicos de entrevistas que utilizan los investigadores, entre ellos los judiciales.

Los códigos lingüísticos son en esencia la lengua o el idioma, constituido por signos o reglas propios de la gramática de cada habla. Dichos signos son utilizados en el Sistema de Etiquetado para registrar los elementos verbales del discurso. Es así como, además del discurso emitido por el entrevistado, en el que se registran todas las palabras expresadas por él o los sujetos que intervienen, se tienen en cuenta aspectos como el tema tratado, el lugar donde fue realizada la entrevista, la fecha y la duración, entre otras. Veamos la siguiente tabla en la que se explica cada categoría.

REGISTRO DE PROYECTO Ficha Técnica (MOREIRO, 2000, p. 326).	
Código o signatura	Referencia dada al documento en el momento de la sistematización, según el fondo documental que se le asigne. Dicho código estará conformado por la fecha de realización de la entrevista, el tipo de temática, el número consecutivo de la entrevista y el código del investigador.
Autor o entrevistador	Autor del documento: director – realizador.
Situación comunicativa	Tema: Temas o materias que representan el contenido del documento (Ver Sistema de Clasificación Decimal Dewey, 2000).
	Propósito: objetivo previsto al momento de entablar una comunicación.
Voz Principal	Nombre de la persona que interviene en la grabación como protagonista.
Genero	Masculino o femenino.

REGISTRO DE PROYECTO Ficha Técnica (MOREIRO, 2000, p. 326).	
Edad	= 25, 26-55, > 55.
Idioma	Lengua materna o dialecto del hablante
Nivel de estudios	Analfabetos, primarios, secundarios, medios, superiores.
Nivel sociocultural	Alto, medio, bajo.
Otras voces	Nombre(s) de otra(s) persona(s) que interviene en la grabación cuya voz se escucha en el documento. En esta categoría deben ir definidas las iniciales del nombre u oficio que caracterizaran el informante, todo con el objetivo de abreviar al momento de la transcripción.
Lugar de la grabación	País, departamento o estado, distrito, municipio o cantón, otros.
Fecha de la grabación	Datos exactos del día en que fue grabada la entrevista.
Duración de la grabación	Tiempo de la entrevista registrado en horas, minutos y segundos. El sistema le permitirá al investigador visualizar el tiempo de la grabación mientras transcribe y etiqueta, con el objetivo de hacer un registro fiel de los momentos de la grabación.
Soporte de la grabación	Tipo de soporte, tipo de grabación (mono, estéreo, velocidad de reproducción).
Resumen	Dado que el documento sonoro no es directamente accesible sino a través de su reproducción y audición, un resumen que recoja lo esencial del contenido es de gran ayuda para la recuperación y elección del documento buscado.

**Cuadro 1** – Registro de proyecto: ficha técnica.

La utilidad de estas categorías radica en la funcionalidad de los descriptores, que ubican al investigador en el contexto de la acción comunicativa y permiten almacenar la información con el rigor y detalle que merece la fuente. La descripción formal de los documentos es un proceso importante en las llamadas Ciencias de la Documentación, porque permite administrar mejor la información documental y construir bases de datos electrónicas de fácil reconocimiento. Los analistas del discurso y los académicos que buscamos información en fuentes primarias y secundarias reconocemos el valor de la exactitud en el procesamiento de los datos y sabemos que un documento mal clasificado significa un obstáculo para la investigación.

Ahora, si bien los signos lingüísticos operan en niveles como el fónico, el léxico, el semántico y el sintáctico, su objetivo fundamental es

evidenciar, o descubrir, los problemas que el lenguaje plantea como medio de relación social. Por su parte, los signos paralingüísticos incluyen recursos prosódicos estrechamente relacionados con los signos lingüísticos, pero que apuntan más a la interpretación de aspectos relacionados con la curva melódica de la entonación. Veamos las siguientes tablas que explican mediante signos cada uno de los códigos que necesitaremos para la transcripción:

Signos de transcripción relativos a la prosodia (TUSÓN VALLS, 2003, p. 100-101).	
¿?	Entonación interrogativa
¡!	Entonación exclamativa
<	Tono ascendente
>	Tono descendente
=	Tono sostenido o continuo
...-	Corte abrupto en medio de una palabra
I	Pausa breve
II	Pausa mediana
<...>	Pausa larga, también <pausa> o <09>, indicando el número de segundos que tarda en hablar
†	Tono agudo
‡	Tono grave
Ac	Ritmo acelerado
Le	Ritmo lento
MAYÚS	Mayor énfasis: la palabra será escrita con MAYÚSCULA, con el objetivo de señalar el tono enfático
Aaa	Alargamiento vocálico
Nnn	Alargamiento consonántico
P	Piano (dicho en voz baja)
Pp	Pianissimo (dicho en voz muy baja)
F	Forte (dicho en voz más alta)
Ff	Fortissimo (dicho en voz muy alta)

**Cuadro 2** – Signos de transcripción relativos a la prosodia.

Signos de transcripción relativos a los turnos de palabras (TUSÓN VALLS, 2003, p 101).	
= =	Al principio de un turno para indicar que no ha habido pausa después del turno anterior
=.....=	Solapamiento en dos turnos

**Cuadro 3** – Signos de transcripción relativos a los turnos de palabras.

Según Pierre Guiraud, citado por Niño Rojas (2003, p. 58), existen tres tipos de códigos del lenguaje que facilitan la interpretación del lenguaje verbal. Ellos son: los relevos, los sustitutos y los auxiliares.

Los códigos de relevos son aquellos que no cambian el significado del lenguaje, pero que alteran la forma de representarlos. Ejemplos de estos códigos son el Sistema Braille, el Morse y el lenguaje de los sordomudos, entre otros, los cuales permiten recodificar con bastante fidelidad el lenguaje fónico, introduciendo peculiaridades lingüísticas muy significativas propias de la lengua escrita.

Por su parte, los códigos sustitutos del lenguaje son aquellos que no representan el lenguaje, sino que pretenden sustituirlo. Un claro ejemplo de este tipo de escritos lo encontramos en los ideogramas chinos, donde los caracteres no representan los signos de la lengua, sino ideas y conceptos (NIÑO ROJAS, 2003, p. 58). También hacen parte de ella los jeroglíficos, las pinturas y las siglas.

Finalmente, tenemos los códigos auxiliares del lenguaje, que acompañan la significación de las palabras según sea la entonación y la expresividad corporal. Este último factor es comúnmente conocido como código kinésico y proxémico, e incluye aspectos como los gestos, el movimiento de las manos, la mirada, la utilización del medio ambiente, etc. Veamos la siguiente tabla en la que aparecen signos con las características antes descritas, que obviamente deben ser advertidos por el investigador.

Otros Signos (TUSÓN VALLS, 2003, p. 101).	
[no léxicos]	Fenómenos no léxicos, tanto vocales como no vocales, p. e. [risas], [A mirando a B], [llanto], [duda] ...
(ininteligibles)	Palabras ininteligibles o dudosas que deben ser escritas entre paréntesis tal cual suenan.

**Cuadro 4** – Otros signos.

A esta última categoría, kinésica y proxémica, la denominamos código extralingüístico. Tiene que ver con la capacidad de sentir, nivel sensorial, y está relacionada estrechamente con el contexto comunicativo y con aspectos como la cultura y la ideología del hablante. Los códigos extralingüísticos nacen de la experiencia objetiva, subjetiva y cultural del ser humano, y tiene como propósito significar la relación entre los hombres. Todos estos códigos tienen en común su carácter eminentemente semiótico.

La semiosis incluye todas las formas de creación de significado – las imágenes, el lenguaje corporal y también el lenguaje –. Podemos entender la vida social como una serie de redes interconectadas de prácticas sociales de diferentes tipos. Todas las prácticas son prácticas de producción que constituyen los escenarios en los que se produce la vida social, ya sea ésta económica, política, cultural o de carácter cotidiano. (FAIRCLOUGH, 2003, p. 179-180).



Socialmente, dichos códigos pueden ser catalogados de la siguiente forma (NIÑO ROJAS, 2003, p. 61):

- a) signos de identidad: los registros civiles, las insignias, los banderines, símbolos patrios, uniformes, nombres y sobrenombres, maquillajes y tatuajes;
- b) signos de cortesía: formulas y saludos, tonos especiales y gestos convencionales (inclinación de cabeza, estrechón de manos, abrazo, beso, etcétera);
- c) costumbres, hábitos y utensilios: alimentos, comercio, muebles, arquitectura, instrumentos y maquinaria;
- d) ritos y reuniones: los ritos religiosos, protocolos y etiquetas, fiestas, regalos, noviazgo, matrimonio, clan, tribu, gremio y demás signos relativos a las comunidades;
- e) modas: como el vestido, danzas y costumbres predominantes;
- f) juegos y diversiones: deportes, espectáculos, competencias;
- g) patrimonio político y cultural: leyes, valores, medios de comunicación, ciencias y artes.

Estos códigos no necesariamente tienen una intencionalidad específica, sino que la determina el contexto social, es decir, la significación que le otorgan las personas que habitan una sociedad y que, de hecho, hacen parte de ella. Es así como de este espectro de representaciones semióticas hacen parte las expresiones dialectales, los artificios de uso colectivo y personal, los rituales y todo aquello que distingue y asemeja una sociedad del resto.

## 5 Software AVOCES

La aplicación permite a los investigadores transcribir discursos orales en un entorno integrado de video y editor de texto, con opciones como análisis de frecuencia de las palabras y búsqueda por contexto.

Tal y como se aprecia en la imagen, el sistema contiene dos interfaces, una para el administrador y otra para el investigador. De un lado está el administrador, quien se encarga de registrar y controlar las acciones de los investigadores, a la vez que examina el ingreso adecuado de los discursos orales. De otro lado está el investigador, quien posterior al registro efectuado por el administrador debe etiquetar y la transcribir las entrevistas, conservando todas las especificaciones dadas por el sistema.



**Print Screen 1** – Interfaces para el administrador y para el investigador.

El rigor de la fuente oral exige que no se omita ninguna de estas consideraciones, con el ánimo de lograr una buena preservación de la memoria oral. Existen dos opciones para que los investigadores registren los proyectos: una general para discursos orales de investigaciones del área socio humanística, y otra específica del área del Derecho Judicial.

**AVOCES**

**Registro de entrevista:**

**Datos de la entrevista**

Discurso Oral  
 Sistema Penal Acusatorio

Código de Proyecto: 12      Género: Masculino  
 Código de Entrevista: 123      Edad: 26-55  
 País: Sistema Penal      Idioma: Español  
 Tema: 410 Lingüística      Nivel de Estudio: Primario  
 Propósito: Análisis del Discurso      Nivel Sociocultural: Bajo  
 Voc Principal: Juan Esteban Arango      Entidades Involucradas: Fiscalía

**Datos de la grabación**

Lugar: Medellín      Resumen: funciones. En esta audiencia se evidencia varios elementos, tales como la capacidad de dirección del proceso por parte del juez, para llevar el orden de la audiencia, además de admitir testimonios, retulaciones y alegatos dentro del curso de la misma.  
 Fecha: Viernes, 06 de Enero de 2006  
 Duración: 00:20:25 (hh:mm:ss)  
 Soporte: DVD

(\*) Campos con Información Obligatoria  
Cuándo no se registre información en los campos no obligatorios, aparecerá "Sin Información" en el registro.

Guardar Cancelar

**Print Screen 2 – Opciones para los investigadores.**

Tal y como lo mencionamos, en caso de que la intervención oral sea una audiencia pública, el investigador deberá registrar datos legales en el cuadro que veremos a continuación:

**AVOCES**

**Registro de entrevista:**

**Descripción legal**

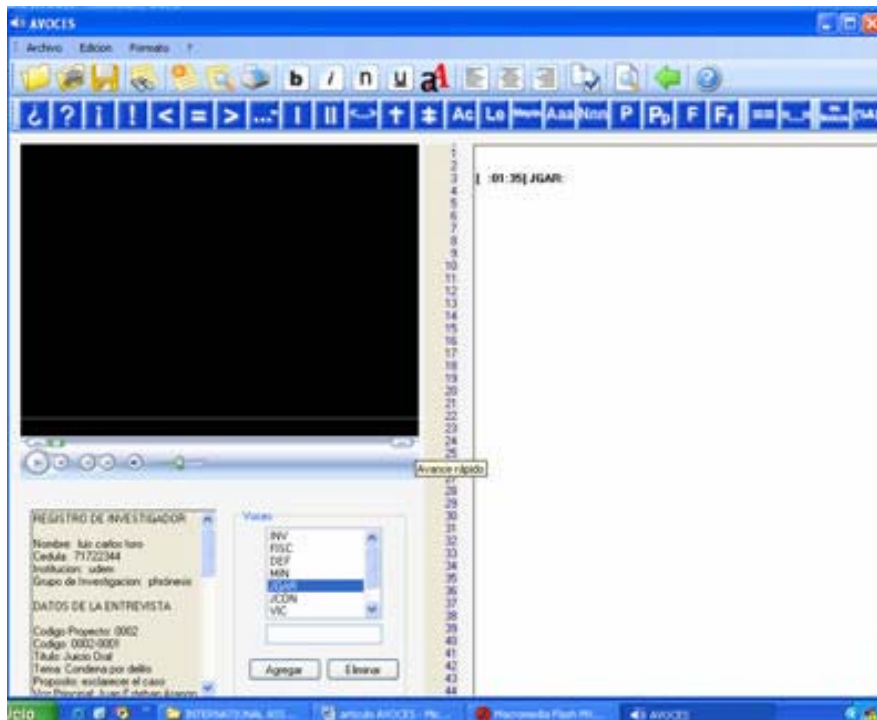
Entrevista: 123  
 Hoja de Vida: 12345  
 Registro Caso: 50016000205200500178  
 Registro Expediente: 050014086001  
 Juzgado: Primero de Medellín  
 Distrito: Medellín - Antioquia  
 Folio: 12345

Todos los campos son obligatorios.

Guardar Cancelar

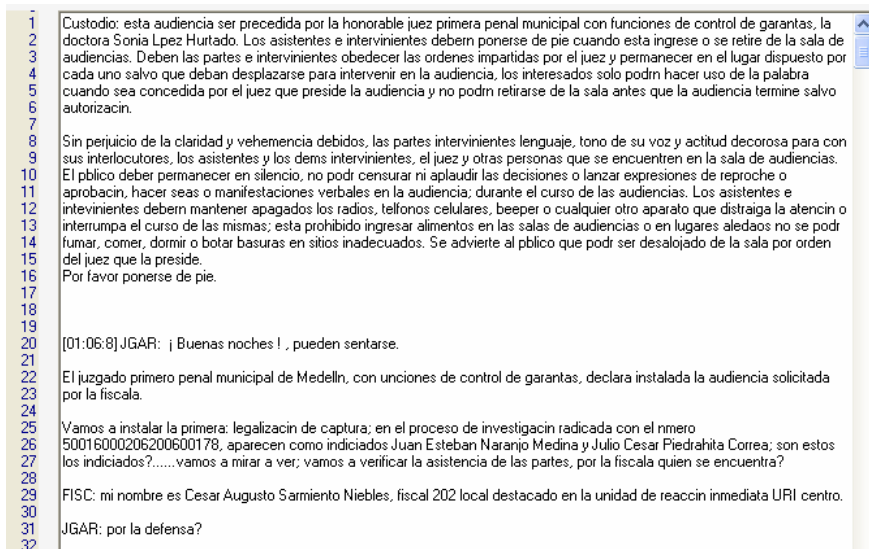
**Print Screen 3 – Registro de datos legales.**

Una vez registrado el proyecto y la entrevista, ingresamos al editor de texto. El editor es quizás la parte más dinámica del sistema, porque incorpora las entrevistas realizadas en video o audio y la transliteración de las mismas. Dicho editor le permite al investigador realizar una serie de marcas prosódicas, muy propias del discurso hablado, que servirán posteriormente para el análisis que se quiere realizar.



Print Screen 4 – Editor de texto.

A la hora de transcribir el texto el investigador debe tener en cuenta las etiquetas descritas en el apartado cuatro, denominadas Signos de Transcripción. Tal y como se aprecia en la siguiente grafica, dicha información está debidamente cronometrada y numerada, es decir, identificando el minuto aproximado en el que interviene los hablantes y la línea en la que fue transcrita. Dicha sistematización se hace con el fin de facilitar la búsqueda de la información en los cassettes y en el texto.

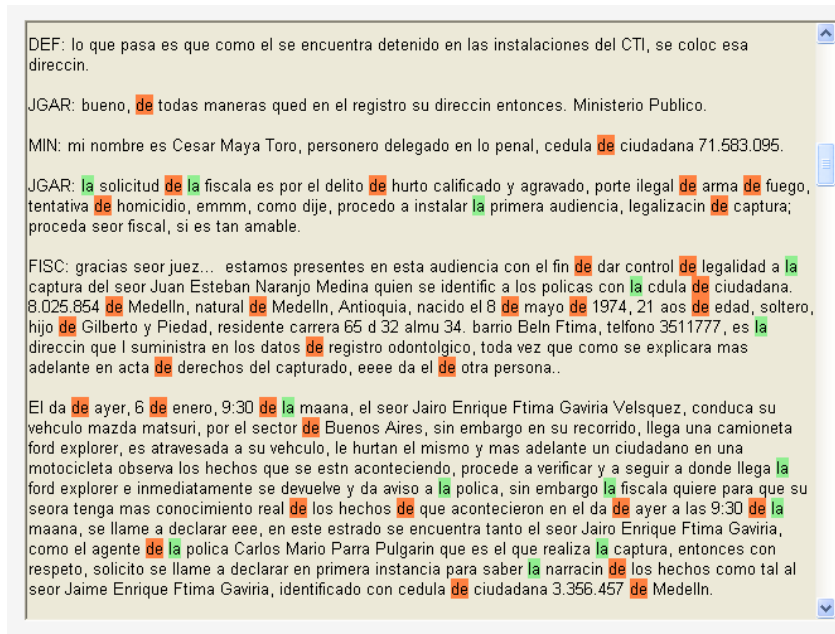


**Print Screen 5 – Entrevista transcrita.**

Los análisis que se derivan de las entrevistas transcritas son el resultado de una lectura cuidadosa, tal y como lo vemos en la siguiente interfaz, que permite al investigador ver la ubicación de las palabras utilizadas en el discurso por orden alfabético, por frecuencia y por número de caracteres. Esto a su vez permite marcar las palabras de interés, y obtener un informe de usos y coincidencias prosódicas del discurso hablado.



**Print Screen 6 – Ubicación de las palabras utilizadas en el discurso por orden alfabético, por frecuencia y por número de caracteres.**



Print Screen 7 – Marcas de las palabras de interes.

Finalmente, un análisis por ubicación permite ver los microcontextos en los que encuentra el concepto de interés para el investigador. Dicho análisis se hace con el objetivo de buscar coincidencias en un discurso o una serie de discursos orales, lo que permite un estudio comparado de fenómenos léxicos y/o fisuras discursivas.



Print Screen 8 – Microcontextos de interes.

## 6 Conclusões

Si bien todavía es necesario que los discursos orales sean transcritos manualmente, es decir, que pasen por el filtro de la textualidad para un mejor reconocimiento, traducción, resumen y clasificación de los mismos, debemos reconocer que el etiquetado constituye un avance significativo en el estudio de la interacción comunicativa, que amplía nuestra perspectiva sobre el tema tratado, al punto de atrevernos a plantear que la marcación de los discursos es la primera fase del análisis.

En este sentido, este sistema permite a los investigadores del área socio humanística precisar aspectos relativos al lenguaje, que sirven como soporte para consolidar un proceso encaminado al análisis del discurso oral en Colombia. Además, dicho avance tecnológico permite conservar todo tipo de manifestaciones orales en soporte magnético o digital, lo que conlleva necesariamente a una redefinición en la forma de contar el pasado y procesar la información.

Con todo, si bien el etiquetado contribuye un enriquecimiento de los documentos orales, es necesario entender que por más esfuerzos teóricos y metodológicos que se emprendan en la selección y definición de categorías, cada usuario marcará los discursos de acuerdo al nivel de conocimientos y de observación que posea, lo que constituye un reto en el proceso de clasificación del lenguaje.

*Recebido em outubro de 2007.*

*Aprovado em junho de 2009.*

*AVOCES: System to label oral discourses*

### **Abstract**

The oral communication is natural to human beings and it makes part of individual as a member of specie. It is produced in and with a body, whose respiratory and deglutition systems organs are taken advantage of. Also, eyes movement, different facial expressions, and other corporal movements are an important part of the orality, such as "vocalizations" (vocal sounds, but not linguistics) and other "noises". Nevertheless, a lecture, a speech, or a judgment for instance, even though all of them are oral manifestations, they request a high degree of preparation, including the writing use (a scratch, some notes, some proceedings, etc). Controlling these ways of speaking is not easy. Thus, the development of the oral communication performance is also part of linguistics education. Considering what was said, this paper aims to present decisions about the conceptions of helpful categories in order to analyse oral discourses, especially the ones related to labeling and transcription. These orientations are followed by theoretical assumptions that support them, besides a traced algorithm that organizes the systematization processes and the storage of these kind of sources.

**Keywords:** Oral Discourse Analysis. Linguistics Technology. Labeling and Transcription.

## Fuentes Bibliográficas

BAZALAR DÍAZ, Jacqueline M. Características y diferencias entre el código lingüístico oral y el código lingüístico escrito. In: **El lenguaje**. 2006. Disponible en: <<http://www.monografias.com/trabajos32/lenguaje/lenguaje.shtml#codigo>>. Acceso en: 4 sept. 2006.

BLANCHE-BENVENISTE, Claire. **Estudios lingüísticos sobre la relación entre oralidad y escritura**. Barcelona: Gedisa, 1998.

BRIGGS, Asa; BURKE, Peter. La revolución de la imprenta en su contexto. In: BRIGGS, Asa. **De Gutenberg a Internet: una historia social de los medios de comunicación**. Madrid: Taurus, 2002.

BRIZ, Antonio y Grupo Val.Es.Co. **¿Cómo se comenta un texto coloquial?** Barcelona: Ariel Practicum, 2002.

CIVALLERO, Edgardo. **Tradición Oral: herramientas y experiencias sobre oralidad, su revitalización, recolección, gestión y difusión**. Disponible en: <<http://www.tradicionoral.blogspot.com/>>. Acceso en: 1 ago. 2006.

CALSAMIGLIA, Helena; TUSÓN, Amparo. **Las cosas del decir: manual de Análisis del Discurso**. Barcelona: Ariel, 2002.

DEBRAY, Régis. **Trasmitir**. Buenos Aires: Manatíal, 1997.

DEBRAY, Régis. **Introducción a la mediología**. Barcelona: Paidós, 2001.

DU BOIS, John J. Transcription design principles for spoken discourse research. **Pragmatics**, n. 1, p. 71-106, 1991.

DÜRKEIM, Émile. **Las reglas del método sociológico**. Madrid: Morata, 1993.

EUROMAP. **Estudio comparativo de la evolución de las tecnologías lingüísticas en Europa**. Resumen en español del informe final del proyecto EUROMAP de Rose Lockwood y Andrew Joscelyne. Copenhagen: Information Society Technologies, Instituto Cervantes. 2003.

FAIRCLOUGH, Norman. El análisis crítico del discurso como método para la investigación en ciencias sociales. In: WODAK, Ruth; MEYER, Michael. (Comp.). **Método de análisis crítico del discurso**. Barcelona: Gedisa, 2003.



FEBVRE, Lucien. **Combates por la historia**. Barcelona: Planeta Agostini, 1993. Colección Obras Maestras del pensamiento contemporâneo, n. 28.

GONZÁLEZ, Luis. El arte de la microhistoria (Ponencia presentada al Primer Encuentro de Historiadores de Provincia, San Luis Potosí, 26 de julio de 1972). Disponible en: <[http://biblioteca.redescolar.ilce.edu.mx/sites/fondo2000/vol1/otra\\_invitacion/html/1.html](http://biblioteca.redescolar.ilce.edu.mx/sites/fondo2000/vol1/otra_invitacion/html/1.html)>. Acceso en: 16 mayo 2006.

GUIRAUD, Pierre. **La Semiología**. México: Siglo XXI, 1971.

GUMPERZ, Jhon J.; BERENZ, Nadine. Transcribing conversational exchanges. **Berkeley Cognitive Science Report**, n. 63, 1990.

JÄGER, Siegfried. **Kritische Diskursanalyse: Eine Einführung**. Münster: Unrast-Verlag, 2004.

HABILIDADES comunicativas del defensor en el juicio oral. Colombia. Unidad de Capacitación, Dirección Nacional de Defensoría Pública. Financiado por: Programa de Fortalecimiento y Acceso a la Justicia.

HALLIDAY, M.A.K. **Spoken and written language**. Oxford: University Press, 1989.

LEVINSON, Stephen C. **Pragmatics**. Cambridge: Cambridge University, 1983.

LYONS, J. **Introducción a la lingüística teórica**. Barcelona: Teide, 1980.

LLISTERRI, J. Transcripción, etiquetado y codificación de corpus orales. **Revista Española de Lingüística Aplicada: Panorama de la Investigación en Lingüística Informática**, p. 53-82, 1999.

LOPEZ MORALES, Humberto. **Las Antillas**. Madrid: Arco Libros, 1994.

MORIN, Violette. **Écriture de presse**. Paris: Mouton, 1969.

MORIN, Violette. **Tratamiento Periodístico de la Información**. Barcelona: ATE, 1974.

MOREIRO, José Antonio. **Manual de documentación informativa**. Madrid: Cátedra, 2000. p. 326.

NIÑO ROJAS, Víctor Miguel. **Competencias en la comunicación**. Bogotá: Eco, 2003.

ONG J. Walter. **Oralidad y escritura: tecnologías de la palabra**. Colombia: Fondo de Cultura Económica, 1994.

PAYRATÓ, Lluís. Transcripción del discurso coloquial. In: **El español coloquial: actas del primer simposio sobre análisis del discurso oral**. Almería: Universidad de Almería, 1995. p. 45-70.

RENKEMA, Jan. **Introducción a los Estudios sobre el Discurso**. Barcelona: Cladema, 1999.

SACKS, Harvey; SCHEGLOFF, Emmanuel A.; JEFFERSON, Gail. A simplest systematics for the organization of turn-taking in conversation. **Language**, n. 50, p. 731-734, 1974.

SCOLLON, Ron; SCOLLON, Suzie Wong. **Discourses in place: language in the material world**. London: Routledge, 2003.

SEARLE, John. **Actos de habla**. Madrid: Cátedra, 1980.

SIMONE, Raffaele. **La tercera fase: formas de saber que estamos perdiendo**. Madrid: Taurus Pensamiento, 2001.

SISTEMA de Clasificación Decimal Dewey. Bogotá: Rojas Eberhard, 2000.

TAN A. H. Text Mining: the State of the Art and the Challenges. In: **PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases**, Beijing, China. 1999.

TORO TAMAYO, Luis Carlos. **Análisis argumentativo, retórico y semiótico de discursos publicitarios en medios impresos: el caso de las tecnologías de comunicación masiva en Colombia**. 2006. Fabio Mejía Fernández (director). Trabajo de grado. 5ª Cohorte Maestría en Lingüística, Facultad de Comunicaciones de la Universidad de Antioquia. Medellín.

TUSÓN, Jesús. **La escritura: una introducción a la cultura alfabética**. Barcelona: Octaedro, 1997.

TUSÓN VALLS, Amparo. **Análisis de la Conversación**. Barcelona: Ariel Practicum, 2003.

VAN DIJK, Teun A. **La Noticia como discurso**: comprensión, estructura y producción de la información. Barcelona: Paidós, 1990.

VAN DIJK, Teun A. **Las estructuras y funciones del discurso**. 7. ed. México: Siglo XXI, 1991.

VAN DIJK, Teun A. **Prensa, racismo y poder**. México: Universidad Iberoamericana, 1995.

VAN DIJK, Teun A. **Racismo y análisis crítico de los medios**. Barcelona: Paidós, 1997.

VAN DIJK, Teun A. **Ideología**: un enfoque multidisciplinario. Barcelona: Gedisa, 1998.

VAN DIJK, Teun A. (Comp.). **El discurso como interacción social**. Barcelona: Gedisa, 2000.

VAN DIJK, Teun A. **Ideología y discurso**: una introducción multidisciplinaria. Barcelona: Ariel, 2003. Informes

VAN EEMEREN, Frans H.; VAN EEMEREN, F. H.; GROOTENDORST, Rob. **Argumentation**: analysis, evaluation, presentation. New Jersey: Lawrence Erlbaum, 2002.

WESTON, Anthony. **Back to earth**: tomorrow's environmentalism. Philadelphia: Temple University Press, 1994.