

# **Universidad Católica “Nuestra Señora de la Asunción”**

**Facultad de Ciencias y Tecnología (CYT)**

**Ingeniería Informática**

**Teoría y Aplicaciones de la Informática 2**

**Tema: Interfaces de Lenguajes Natural**

**Rodrigo Villalba Zayas**

**Año 2007**



## Contenido

- Introducción pag. 3
- Procesamiento del lenguaje Natural pag. 3
- Breve Historia pag. 4
- Dificultades en el Procesamiento de Lenguajes Naturales pag. 4
- Aplicaciones pag. 5
  - Síntesis de Voz pag. 5
  - Reconocimiento del Habla pag. 6
  - Reconocimiento de Caracteres Ópticos pag 7
  - Generación de Lenguajes Naturales pag. 8
  - Traducción Automática pag. 8
  - Búsqueda de Respuestas pag. 9
  - Recuperación de Información pag. 10
  - Extracción de la Información pag. 10
- Bibliografía pag 10
- Anexos pag 11

## Introducción:

El lenguaje natural es uno de los muchos estilos de interfaces que pueden ser utilizados en un dialogo entre una computadora y un humano. Como veremos es muy difícil dirigir las operaciones de una maquina usando el mismo lenguaje que usamos en el día a día. Campos en la ciencia de la computación estipulan que el lenguaje natural esta lejos de ser una alternativa atractiva como parecería al principio. La literatura tiende a enfocarse a los problemas potenciales que esta presentaría. Unos de esos problemas sería la ambigüedad que presentan los lenguajes naturales (en contrapartida de los lenguajes formales usados para programar computadoras que son fácilmente traducibles a sentencias de procesador). Implementaciones exitosas han sido caracterizadas por ser suficientemente restringidos en el análisis sintáctico y semántico de dicho lenguaje poniendo en duda el estatus de “Lenguaje Natural”.

Usualmente el grado de ambigüedad de los lenguajes naturales es considerado demasiado extremo para ser usado efectivamente como un “Estilo de Interfaz” Cuando los sistemas usan restricciones como estructuras para limitar la ambigüedad, es asumido que el usuario deberá aprender que estructuras son aceptables haciendo que “el lenguaje natural” no sea más útil o aprendible que un lenguaje de comandos formal. A pesar de la falta de utilidad predicha por algunos entendidos del campo, algunos sistemas de lenguaje restringido mantienen las mayorías de las ventajas de las interfaces de lenguajes natural. Usualmente muchos problemas no son debido a la ambigüedad sino a la expectativa excesiva en la capacidad del sistema de parte del usuario o el conocimiento del mundo que no esta reflejado en la base de conocimientos de la computadora.

## Procesamiento del lenguaje Natural

El **Procesamiento de Lenguajes Naturales** (PLN) es una subdisciplina de la Inteligencia Artificial y de la computación lingüística. El PLN se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio de lenguajes naturales. El PLN no trata de la comunicación por medio de lenguajes naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente, que se puedan realizar por medio de programas que ejecuten o simulen la comunicación. Sistemas de generación de lenguajes naturales convierten información de bases de datos almacenadas en computadoras en lenguaje natural humano (escrito o hablado) y sistemas de comprensión de lenguajes naturales convierten lenguaje natural humano (escrito o hablado) en una representación formal fácilmente manipulables por programas de computador.

En teoría el procesamiento del lenguaje natural es un método muy atractivo en la interacción humana-computador. Los primeros sistemas como SHRDLU trabajaban en mundos muy restringidos con vocabularios muy restringidos con gran éxito, llevando a los investigadores a un optimismo excesivo que fue rápidamente perdido cuando dichos sistemas eran extendidos a situaciones más realistas con ambigüedad y complejidad del mundo real.

## Breve Historia

El Procesamiento del Lenguaje Natural (PLN) es una de las piedras angulares tempranas de la inteligencia artificial (IA). La Traducción automática, por ejemplo, nació a finales de la década de los cuarenta, antes de que se acuñara la propia expresión «Inteligencia Artificial». No obstante, el PLN ha desempeñado múltiples papeles en el contexto de la IA, y su importancia dentro de este campo ha crecido y decrecido a consecuencia de cambios tecnológicos y científicos. Los primeros intentos de traducir textos por ordenador a finales de los cuarenta y durante los cincuenta fracasaron debido a la escasa potencia de los ordenadores y a la escasa sofisticación lingüística. Sin embargo, los esfuerzos realizados en las décadas de los sesenta y los setenta para producir interfaces en lenguaje natural para bases de datos y otras aplicaciones informáticas obtuvieron un cierto grado significativo de éxito. La década de los ochenta y el principio de la de los noventa han visto resurgir la investigación en el terreno de la Traducción Automática.

## Dificultades en el procesamiento de lenguajes naturales

### - Ambigüedad

El lenguaje natural es inherentemente ambiguo a diferentes niveles:

- A nivel léxico, una misma palabra puede tener varios significados, y la selección del apropiado se debe deducir a partir del contexto oracional o conocimiento básico. Muchas investigaciones en el campo del procesamiento de lenguajes naturales han estudiado métodos de resolver las ambigüedades léxicas mediante diccionarios, gramáticas, bases de conocimiento y correlaciones estadísticas.
- Ambigüedad sintáctica: La gramática para lenguajes naturales es ambigua (a menudo hay múltiples árboles sintácticos para una sentencia dada). Elegir la más apropiada requiere información contextual y semántica.
- A nivel pragmático, una oración, a menudo, no significa lo que realmente se está diciendo. Elementos tales como la ironía tienen un papel importante en la interpretación del mensaje.

Para resolver estos tipos de ambigüedades y otros, el problema central en el PLN es la traducción de entradas en lenguaje natural a una representación interna sin ambigüedad, como árboles de análisis.

## Detección de separación entre las palabras

En la lengua hablada no se suelen hacer pausas entre palabra y palabra. El lugar en el que se debe separar las palabras a menudo depende de cuál es la posibilidad que mantenga un sentido lógico tanto gramatical como contextual. En la lengua escrita, idiomas como el chino mandarín tampoco tienen separaciones entre las palabras.

## Recepción imperfecta de datos

Acentos extranjeros, regionalismos o dificultades en la producción del habla, errores de mecanografiado o expresiones no gramaticales, errores en la lectura de textos mediante OCR

# Aplicaciones

Las principales tareas de trabajo en el PLN son:

- Síntesis de voz
- Reconocimiento del habla
- Reconocimiento de Caracteres Ópticos
- Generación de lenguajes naturales
- Traducción automática
- Recuperación de la información
- Respuesta a preguntas
- Extracción de la información
- Resumen Automático

## Síntesis de voz

La **voz sintética** es una voz artificial (no pregrabada), generada mediante un proceso de sinterización del habla.

La **Síntesis de habla** es la producción artificial de habla humana. Un sistema usado con este propósito recibe el nombre de **sintetizador de habla** y puede llevarse a cabo en software o en hardware. La síntesis de voz se llama a menudo en inglés **text-to-speech (TTS)**, en referencia a su capacidad de convertir texto en habla. Sin embargo, hay sistemas que en lugar de producir voz a partir de texto lo hacen a partir de representación lingüística simbólica en habla. Sintetizadores del habla pueden ser creados concatenando piezas de voz grabada que son almacenadas en una base de datos. Para dominios específicos el almacenamiento de palabras enteras permite una salida de alto rendimiento y calidad.

La *calidad* de una voz sintética vendrá dada por:

- Su *inteligibilidad*: ¿con qué facilidad/dificultad es entendida?.
- Su *naturalidad*: ¿en qué medida se asemeja a la voz *real* de un humano?

Un sintetizador del habla inteligible permite a personas con problemas visuales o de lectura escuchar trabajos escritos en una computadora personal.

## Sistemas Operativos de computadoras con sintetizadores del habla

\* Apple: El primer sintetizador del habla integrado a un sistema operativo fue MacInTalk de Apple en 1984. En la década de los 90 Apple con la aparición de PowerPc más rápidas fueron incluyendo sintetizadores con calidad de voz mas alta.

\* Microsoft Windows: Windows 2000 incluyo un programa sintetizador del habla llamado Narrator directamente disponibles a usuarios. Microsoft Speech Server es un paquete completo para sintetizadores y reconocedores del habla para aplicaciones comerciales como call centers.

### \* Otros

- Las computadoras personales de Texas Instruments TI-99/4 and TI-99/4A eran capaces de recitar palabras completas y frases. Usando un muy popular periférico sintetizador del habla. TI usaba un codec para empotrar frases habladas completas dentro de aplicaciones como video juegos.
- Sistemas que operan sobre GNU/Linux son varios e incluyen programas open source como Festival Speech Synthesis System.
- Productos comerciales para Pocket Pcs y Smartphones incluyen un programa llamado Smart Read Mobile

## Lenguajes de Mercado de Sintetizadores del Habla

Un gran numero de lenguajes de marcado han sido desarrollados para la representación de texto a hablar en un formato XML compliant. El mas reciente es Speech Synthesis Markup Language (SSML) que fue una recomendación W3C en el 2004.

### Programas Específicos

- Apple PlainTalk
- Festival Speech Synthesis System
- FreeTTS
- Praat
- Software Automatic Mouth
- Text2Speech
- Vocoder
- IVONA TTS
- Kurzweil 1000 and Kurzweil 3000

## Reconocimiento del habla

El **Reconocimiento del Habla** es el proceso de convertir una señal del habla a una secuencia de palabras en la forma de datos digital por medio de un algoritmo implementado como un programa de computadora. Aplicaciones del reconocimiento del habla que han emergido en los últimos años incluyen discado por medio de la voz (Ej. "llamar a casa"), entrada de datos simples (Ej. Ingresar un número de tarjeta de crédito), preparación de documentos estructurados (Ej. Un reporte de radiología).

El Reconocimiento de voz es un proceso relacionado que trata de identificar la persona que esta hablando, sin importarle que esta diciendo (opuesto al reconocedor del habla).

### Usos y aplicaciones

Aunque en teoría cualquier tarea en la que se interactúe con un ordenador puede utilizar el reconocimiento del habla, actualmente las siguientes aplicaciones son las más comunes:

- **Dictado automático:** El dictado automático es, en la actualidad, el uso más común de las tecnologías de reconocimiento del habla. En algunos casos, como en el dictado de recetas médicas y diagnósticos o el dictado de textos legales, se usan corpus especiales para incrementar la precisión del sistema.
- **Control por comandos:** Los sistemas de reconocimiento de habla diseñados para dar órdenes a un computador (p.e. "Abrir Firefox", "cerrar ventana") se llaman Control por comandos. Estos sistemas reconocen un vocabulario muy reducido, lo que incrementa su rendimiento.
- **Telefonía:** Algunos sistemas PBX permiten a los usuarios ejecutar comandos mediante el habla, en lugar de pulsar tonos. En muchos casos se pide al usuario que diga un número para navegar un menú.
- **Sistemas portátiles:** Los sistemas portátiles de pequeño tamaño, como los relojes o los teléfonos móviles, tienen unas restricciones muy concretas de tamaño y forma, así que el habla es una solución natural para introducir datos en estos dispositivos.
- **Sistemas diseñados para discapacitados:** Los sistemas de reconocimiento del habla pueden ser útiles para personas con discapacidades que les impidan teclear con fluidez, así como para personas con problemas auditivos, que pueden usarlos para obtener texto escrito a partir de habla. Esto permitiría, por ejemplo, que los aquejados de sordera pudieran recibir llamadas telefónicas.

### Curiosidades

- Los investigadores del grupo de reconocimiento del habla de Apple solían llevar una camiseta en la que se podía leer *I helped Apple wreck a nice beach* (ayudé a

Apple a estropear una buena playa), cuya pronunciación es idéntica a *I helped Apple recognize speech* (ayudé a Apple a reconocer habla). Esta broma ilustra la dificultad de desambiguar cadenas fonéticas.

## Reconocimiento de Caracteres Ópticos

El reconocimiento de caracteres ópticos (OCR) es la traducción mecánica o electrónica de imágenes escritas a mano o texto impreso (usualmente capturado por un scanner) en texto de maquina editable.

OCR es un campo de investigación en las áreas de Reconocimiento de Patrones, Inteligencia Artificial y Óptica Computarizada. El Reconocimiento de caracteres ópticos (usando técnicas ópticas como espejos y lentes) y el Reconocimiento de caracteres digitales (usando scanner y algoritmo de computadoras) fueron originalmente considerados campos separados. Porque muy pocas aplicaciones que usan técnicas ópticas sobreviven, el termino OCR ha sido expandido para incluir también el procesamiento de imágenes digitales.

Los primeros sistemas requerían entrenamiento (la provisión de ejemplos conocidos de cada carácter) para poder leer una fuente específica. Sistemas inteligentes con un alto grado de precisión en el reconocimiento para muchas fuentes son ahora muy comunes.

## OCR software

- ABBYY FineReader OCR
- GOCR
- Falcon32
- IPStudio
- Microsoft Office Document Imaging
- NovoDynamics VERUS
- Ocrad
- Ocropus
- OmniPage
- Readiris
- ReadSoft
- SmartScore
- Tesseract (software)
- TopSoft TopOCR

## Generación de lenguajes naturales

La **Generación de Lenguajes Naturales** (GLN) es el proceso de la construcción de un texto en lenguaje natural para la comunicación con fines específicos. El principal énfasis de la generación de lenguajes naturales no es sólo el facilitar el uso del



ordenador sino también el desarrollar una teoría computacional de la capacidad del lenguaje humano.

Un generador de lenguaje natural típicamente tiene acceso a un gran conjunto de conocimiento del cual ha de seleccionar información para presentar a los usuarios en varias formas. El generar texto es, pues, un problema de toma de decisiones con múltiples restricciones: de conocimiento proposicional, de herramientas lingüísticas disponibles, de los objetivos de la comunicación del usuario a quien se dirige el texto, y de la situación y del discurso pasado. Se trata de identificar los factores involucrados en este proceso y de determinar la mejor forma de representar estos factores y sus dependencias.

Para generar un texto, un sistema debe escoger cierta información de la base de conocimiento, decidir cómo organizarla, y determinar cómo producir el texto en lenguaje natural, lo cual incluye el decidir acerca de la entrada del léxico y de las estructuras sintácticas.

## Traducción automática

La **traducción automática** (TA), también llamada **MT** (del inglés *Machine Translation*), es un área de la lingüística computacional que investiga el uso de software para traducir texto o habla de un lenguaje natural a otro. En un nivel básico, la traducción por computadora realiza una substitución simple de las palabras atómicas de un lenguaje natural por las de otro. Por medio del uso de corpora lingüísticos se pueden intentar traducciones más complejas, lo que permite un manejo más apropiado de las diferencias en la tipología lingüístico, el reconocimiento de frases, la traducción de expresiones idiomáticas y el aislamiento de anomalías.

La traducción es hoy en día el principal cuello de botella de la sociedad de la información y su mecanización supone un importante avance frente al problema de la avalancha informativa y la necesidad de la comunicación translingüística

Existen numerosas herramientas de traducción automática. En la Web algunas de las más populares son las siguientes:

- **Systran S.A.** <http://www.systran.co.uk>
- **BabelFish (Systran)** <http://babelfish.altavista.com>
- **Google Language Tools** <http://translate.google.com>

## Recuperación de información

La **recuperación de información**, llamada en inglés *Information retrieval* (IR), es la ciencia de la búsqueda de información en documentos, búsqueda de los mismos documentos, la búsqueda de metadatos que describan documentos, o, también, la búsqueda en bases de datos, ya sea a través de Internet, intranet, para textos, imágenes, sonido o datos de otras características, de manera pertinente y relevante.

Los buscadores, tales como Google, Lycos y Copernic, son algunas de las aplicaciones más populares de la recuperación de información. Básicamente hay que construir un Vocabulario, que es una lista de términos en *lenguaje natural*, un algoritmo que incluya las reglas lógicas de la búsqueda {Tabla de verdad} y una valoración de los resultados o *cantidad de información* lograda o posible. Este *motor de búsqueda* es pues el que permite plantear una pregunta con no menos de dos términos y mostrar los resultados mínimos.

## Búsqueda de respuestas

La **búsqueda de respuestas**, llamado en inglés *Question Answering (QA)* es un tipo de recuperación de la información. Dada una cierta cantidad de documentos (tales como *World Wide Web*), el sistema debería ser capaz de recuperar respuestas a preguntas planteadas en lengua natural. *QA* es observado como un método que requiere una tecnología de Procesamiento de lenguaje natural más compleja que otros tipos de sistemas para la Recuperación de documentos, y, en algunos casos, se le observa como un paso por delante de la tecnología del buscador.

## Extracción de la información

La **Extracción de la información** (de sus siglas en inglés *IE, Information Extraction*) es un tipo de recuperación de la información cuyo objetivo es extraer automáticamente información estructurada o semiestructurada desde documentos legibles por una computadora.

Una aplicación típica de *IE* es el escaneado de una serie de documentos escritos en una lengua natural y rellenar una base de datos con la información extraída. Las tendencias actuales en relación con la *IE* utilizan técnicas de procesamiento de lenguaje natural que se centran en áreas muy restringidas. Por ejemplo, la *Message Understanding Conference (MUC)*, o **Conferencia para la Comprensión de Mensajes** es una competición que se ha centrado en los siguientes aspectos durante los últimos años:

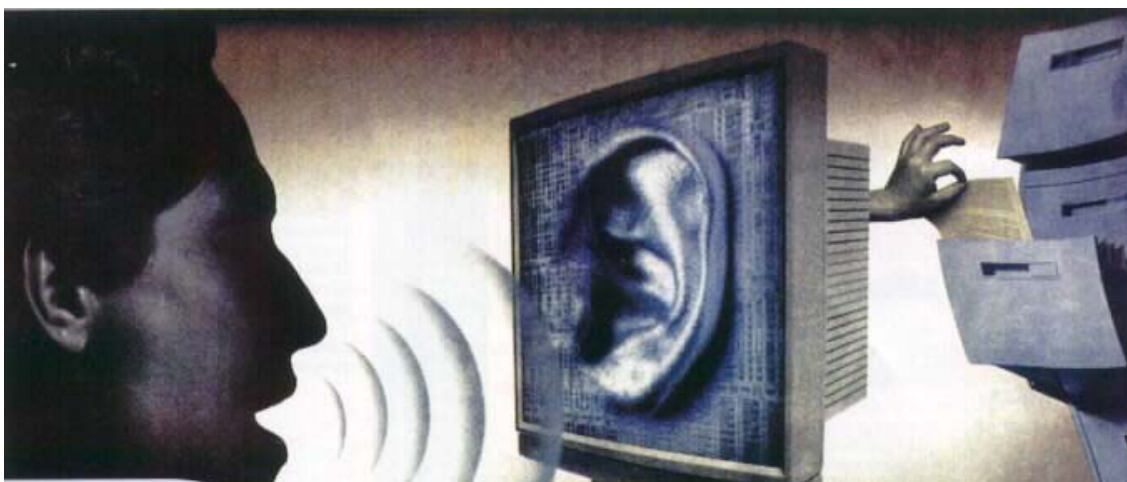
Tareas típicas de la *IE*: Reconocimiento de nombres de personas, organizaciones, lugares, expresiones temporales y ciertas expresiones numéricas.

## Bibliografía

- [http://en.wikipedia.org/wiki/Natural\\_language\\_processing](http://en.wikipedia.org/wiki/Natural_language_processing)
- <http://www.itq.edu.mx/vidatec/espacio/aisc/ARTICULOS/ASCC/articuloASCC.htm>

## Anexos

### **SOFTWARE DE RECONOCIMIENTO DE VOZ**



La precisión del software de Reconocimiento de Voz de las generaciones anteriores alcanzo poco más del 90%, por lo que era cuestionable como herramienta para mejorar la productividad. La buena noticia acerca del software de voz más reciente es que la mayor parte de los productos ofrecen una precisión en el reconocimiento de más del 95% y le ayudan a realizar mayor cantidad de trabajo en menos tiempo, siempre y cuando tenga suficiente velocidad en su computadora, una tarjeta de sonido y un micrófono adecuado y la capacidad de hablar con claridad en todo momento.

El reconocimiento de voz se acerca más a la corriente principal y encuentra su mercado nicho en las comunidades médica y legal, en las que se emplean vocabularios especializados, pero aun no se deseché el teclado y el mouse. Estos programas están diseñados para complementar los medios de captura tradicionales, no para reemplazarlos. Pueden ofrecer un gran incremento en la productividad de los usuarios con habilidades limitadas para mecanografiar. Pero para aquellos que tienen una discapacidad, y que sufren daños constantes por el estrés o simplemente siempre tienen las manos ocupadas estos productos son magníficos.

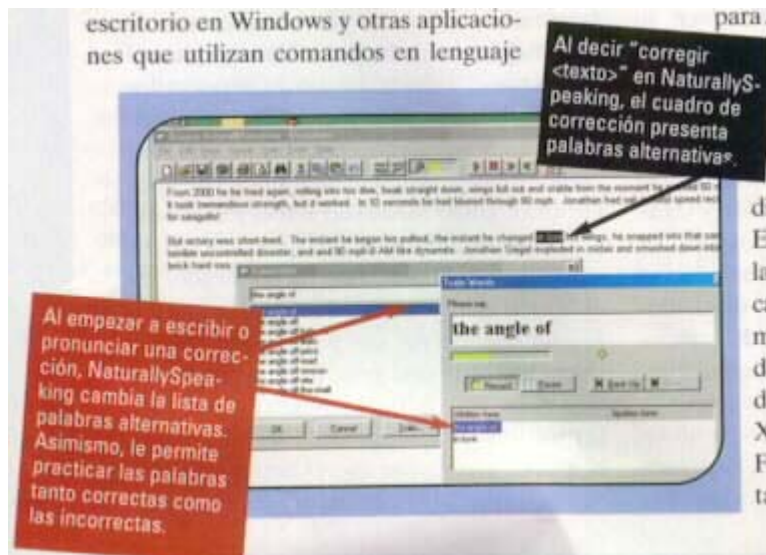
Para esta competencia, los laboratorios de PC Magazine pusieron a prueba, las actualizaciones más recientes de los cuatro productos de reconocimiento de voz: **Dragon Naturally Speaking Preferred 4.0**, **Free Speech 2000 de Philips**, **L & H Voice Xpress Profesional Versión 4.0** y **Via Voice Pro Millenium Edition de IBM**. Todos los fabricantes refinaron sus mecanismos de voz y mejoraron la facilidad de uso e incrementaron el soporte destinado a comandos y controles, así como para el Natural Language Command (NLC) de modo que funcionen en una extensa variedad de aplicaciones, como Microsoft Excel y Outlook.

En los resultados de nuestras pruebas, es evidente que la precisión a mejorado, incluso con menos tiempo de capacitación. En la competencia del año pasado, los evaluadores lograron una precisión promedio de 80 a 90%. Pero para el artículo de este año la precisión inicial va del 91 a 95% en todo el grupo, y aumenta con el tiempo de uso, subiendo hasta 98% en el caso de Via Voice .



La mayor precisión le corresponde a los modelos de lenguaje más inteligentes. Durante nuestras pruebas, observamos una mejor selección de palabras con base en el contexto por parte de Naturally Speaking y Via Voice. Esto no solo produce mayor precisión, también da como resultado la puntuación y aplicación de mayúsculas en forma automática, con ahorro de tiempo al revisar y corregir los dictados. Naturally Speaking, Voice Xpress y Via Voice ahora requieren de unos 10 minutos de lectura para aprenderlos, en lugar de los 30 minutos o más que se necesitaban el año pasado. En la mayor parte de los casos, puede instalar el programa y dictarle con una exactitud muy buena en poco más de 20 minutos. Los tutoriales y las tarjetas de consulta rápida son esenciales para aprender las características del dictado, la navegación y los controles y comandos.

Voice Xpress y Via Voice siempre han tenido un fuerte soporte para controlar su escritorio en Windows y otras aplicaciones que utilizan comandos en lenguaje natural, y ahora amplían ese soporte a los comandos específicos de Microsoft Office 2000, por ejemplo " al enviar como correo electrónico" o "guardar como pagina Web". Todos los productos le permiten crear marcos de voz para insertar texto o dictar en cualquier aplicación de windows.



Todos los productos que participaron en esta competencia ofrecen algunas características Web que se activan con la voz. Natural Web de Naturally Speaking es el más sencillo de usar para navegar en la Web. Le permiten capturar URL con facilidad y seleccionar vínculos para las páginas por medio de la voz. Via Voice indexa cada página en forma dinámica conforme las carga y despliega un número junto a cada vínculo, de modo que solo necesita mencionar el dicho número para mencionar un vínculo. Via Voice ofrece soporte para Jargon de Charter, un vocabulario para los programas de conversiones en Internet. Free Speech actualiza de forma dinámica su Command

Explorer con base en la página web que esta cargada en ese momento para mostrar todos los vínculos activados con la voz, y Via Voice Xpress incluye Web Finder que es una herramienta de metabúsqueda que se activa con la voz. Aún tiene más sentido navegar con la Web con un mouse y un teclado, pero el futuro de la tecnología de reconocimiento de voz podría utilizarse para navegar en la web en dispositivos portátiles más pequeños.

## Dragon Naturally Speaking Preferred 4.0

Precio de lista : \$199 USD. Requiere: Pentium MMX/200 o equivalente; 48 MB en RAM; 200MB de espacio en disco duro; tarjeta de sonido de 16 bits compatible con Creative Labs Sound Blaster; Microsoft Windows 95 , 98 o NT.

Nuestra selección de los editores en esta competencia, Dragon Naturally Preferred 4.0, marca el estándar para la facilidad de uso, desde la instalación directa hasta el formato sencillo y la navegación en el Web activada por la voz.

Con una mejor optimización para los nuevos procesadores y avances en el reconocimiento de palabras y comandos, Naturally Speaking ofrece una exactitud impresionante desde el desempaque. Sólo tiene la desventaja de un soporte limitado para el lenguaje natural en aplicaciones distintas de Microsoft Word. La precisión inicial de Naturally Speaking empató con la de nuestra mención honorífica, el Via Voice, 95%, y aunque este último subió a 98% con capacitación adicional, Naturally Speaking termino en 96%.

Sin embargo, los evaluadores informaron, de manera unánime, haber tenido el menor número de problemas al entender los comandos de voz de Naturally Speaking para la navegación y el formato.

Después de una práctica de 5 minutos, Naturally Speaking tarda alrededor de 2 minutos más en ajustar el modelo de lenguaje de base de acuerdo con su voz y pronunciación. El Vocabulary Builder importa palabras solas, listas de palabras o documentos completos, y le permite practicar todas las palabras o las seleccionadas tan pronto como las importa. Igual que con Via Voice, puede importar documentos para mejorar el reconocimiento.

## Selección de los Editores

### • Dragon NaturallySpeaking Preferred 4.0



*Las pruebas en el área de reconocimiento de voz que realizamos este año mostraron avances en la precisión en todo el grupo, con la calificación más alta (98%) para ViaVoice Pro Millennium Edition de IBM. Pero las calificaciones en exactitud son sólo una parte de la ecuación,*

y para lograrlas tendrá que utilizar las características de navegación, corrección y formato del programa, razón por la cual nuestra Selección de los Editores es para NaturallySpeaking. Con una impresionante calificación máxima en precisión de 96%, NaturallySpeaking queda sólo un poco atrás de ViaVoice. Y después de tomar en cuenta las intuitivas características de edición del programa, esenciales para lograr tal precisión, NaturallySpeaking sobresale entre los demás. Tomemos por ejemplo los comandos para convertir en mayúsculas. ViaVoice requiere que usted diga "capitalize this" antes de una palabra, pero NaturallySpeaking logra el mismo resultado con sólo decir

"cap". Y los evaluadores descubrieron que ViaVoice suele malinterpretar el comando, mientras que NaturallySpeaking trabajó casi sin fallar.

No podemos ignorar las sorprendentes calificaciones en precisión que obtuvo ViaVoice, de modo que le damos una mención honorífica en esta competencia. También nos agradaron las avanzadas capacidades para macros de ViaVoice, que de seguro serán atractivas para los usuarios con más experiencia. ViaVoice es el único programa que le permite aprender de los estilos de los documentos importados, lo cual es muy útil al importar documentos que no son suyos. Con algunas mejoras a la facilidad de uso del producto, ViaVoice sería el programa preferido.

Naturally Speaking tiene atajos intuitivos para el dictado y el formato. Por ejemplo, sólo tiene que decir "all cap" o "cap" para escribir todo en mayúsculas. Los evaluadores encuentran esto más fácil que decir "capitalize" en FreeSpeech y Voice Xpress, o incluso "capitalize this" en Via Voice . Cuando hace una corrección, Naturally Speaking le pide que diga la palabra correcta y la incorrecta, lo cual le ayuda a aprender cómo pronuncia ambas palabras. La corrección de ortografía funciona en forma adecuada, siempre y cuando hable en forma lenta y sin dudar. Naturally Speaking es el único producto en esta competencia que le permite dictar, corregir y practicar sin utilizar las manos.

## FreeSpeech 2000

Precio lista: con audífonos Plantronics, \$99.99 USD; con Philips SpeechMike , \$149.99 USD. Requiere: Pentium/MMX 166 o equivalente; 48 MB en RAM; 100MB de espacio en disco duro; tarjeta de sonido compatible con Sound Blaster; Microsoft Windows 95,98 o NT con SP3.

FreeSpeech 2000 de Philips Speech Products presenta varios avances respecto a su predecesor: soporte para dictado en seis idiomas. Macros de voz y mayor precisión. Y aun incluye el práctico Philips SpeechMike. Pero las bajas calificaciones en precisión, la información limitada y el hecho de tener que cambiar en forma manual entre los modos de dictado y comandos reduce la facilidad de uso del FreeSpeech.

FreeSpeech, con una calificación inicial de 91% en precisión, mejoró en forma sustancial en comparación con su versión anterior. Pero en una categoría en la que un par de puntos porcentuales afecta la utilidad del producto, ésta es una calificación muy baja. FreeSpeech también se colocó en el último lugar en exactitud, a 93%. Probamos el FreeSpeech con SpeechMike y con Plantronics SR1. Obtuvimos una calidad de sonido aceptable durante la activación del audio con SpeechMike, pero sólo después de tres intentos. El programa también carece de una tarjeta de consulta rápida e incluye poca información de los comandos.

El modo de dictado tiene una característica EasyEdit, que destaca las palabras mientras reproduce el dictado, y la característica EasyCorrect del programa le permite interrumpir la reproducción y reemplazar la palabra subrayada con la opción seleccionada de una lista de alternativas. Para utilizar los comandos de los menús y los cuadros de diálogo, o para navegar en un documento, debe cambiar al modo Comandos con la barra de herramientas o un botón especial de SpeechMike. Esto puede ser confuso, sobre todo comparado con los otros programas, que por el momento no tienen modelos. El botón de SpeechMike facilita el proceso, excepto que, cuando lo suelta, el micrófono se apaga y debe pasar en forma manual al modo de Dictado , lo que reduce los beneficios del botón. Para deletrear palabras, debe cambiar al modo Deletrear, y si desea interrumpir la grabación de la voz, recurra al modo Inactivo. Utilizar cuadros de diálogo que requieren una combinación de navegación y dictado puede ser difícil.

Las herramientas para macros de FreeSpeech le permiten insertar varias líneas de texto, ejecutar programas y golpes de teclas y clics del mouse programados en forma previa. FreeSpeech tiene un soporte adecuado para NCL así como para comandos y controles en

todas las aplicaciones. Al igual que Voice Xpress, incluye una lista de comandos de voz en la que es posible buscar por palabras clave.

Los usuarios encontrarán mayor precisión y facilidad de uso en otros programas que participan en esta competencia. Y muchos usuarios se cansarán muy pronto de cambiar en forma manual entre los modos de FreeSpeech.

## **L & H Voice Xpress Professional, Versión 4**

Precio al público: \$150 USD. Requiere: Pentium II; 48 MB en RAM con Microsoft Windows 95 o 98, o 64 MB con Windows NT; 200 MB de espacio en disco duro; tarjeta de sonido de 16 bits compatible con Sound Blaster o micrófono USB.

L & H Voice Xpress Professional ofrece las características básicas que todos esperamos en un programa de reconocimiento de voz: es posible dictar en cualquier aplicación, crear macros de voz y controlar las aplicaciones con comandos de voz. El punto fuerte del programa son sus comandos intuitivos con lenguaje natural para Microsoft Word, Excel y el sencillo procesador de palabras Voice Xpress. Pero sus características de corrección son limitadas; por ejemplo, carecen de reproducción de audio.

La instalación del programa y la lectura de información necesaria que tomó 10 minutos se realizaron sin problemas, pero la instalación del audio fue menos intuitiva. Al probar el volumen del micrófono, no hay indicación que le diga cuándo dejar de leer. Al terminar la instalación, aparece una barra de control en la parte superior de la pantalla, pero no es muy evidente cómo encender el micrófono. El delgado manual incluye pocas pantallas o ilustraciones, por lo que es difícil empezar a trabajar cuando se es nuevo en el reconocimiento de voz. En nuestras pruebas, Voice Xpress quedó en tercer lugar en precisión final, con 94%, detrás de Via Voice y Naturally Speaking. Para el uso casual, quizá sea suficiente, pero para un dictado más formal, querrá una calificación de 95% o más alta.

Voice Xpress es fácil de utilizar para dar formato al texto. Reconoce frases como "convertir en negritas las últimas tres palabras", convertir en negritas las tres palabras anteriores" y muchas otras variaciones del mismo comando. El programa responde con rapidez a lo que usted dice, por lo que la edición y el formato son rápidos y sencillos. Sus comandos para Excel también funcionan en forma adecuada.

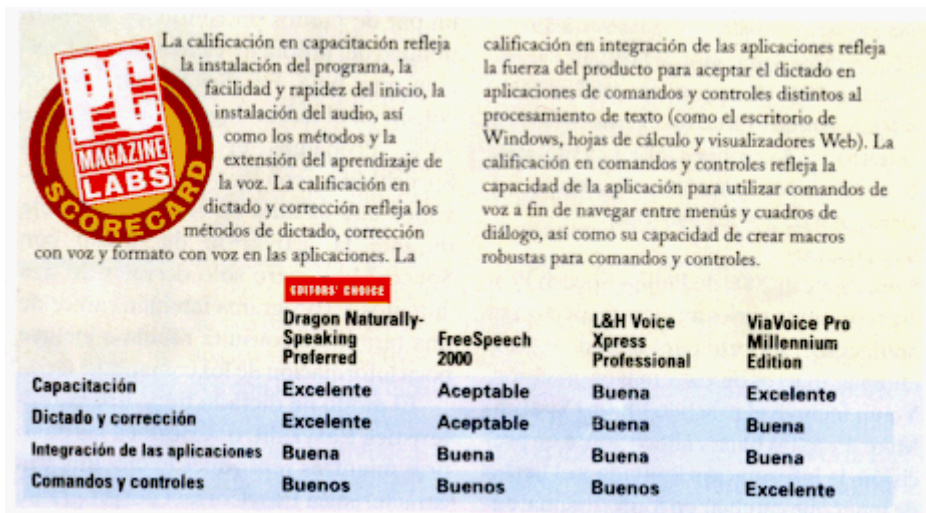
Voice Xpress le permite respaldar sus archivos de voz con facilidad en Iomega Zip Drive u otro lugar seguro. Los usuarios que desean controlar Excel y Word por medio de la voz quizá consideren que Voice Xpress es una opción adecuada.

## **Via Voice Pro Millennium Edition**

Precio al público: \$180 USD. Requiere: Pentium/233 o equivalente; 48 MB en RAM con Microsoft Windows 95 o 98, o 64 MB con Windows NT; 100 MB de espacio en disco duro ; tarjeta de sonido de 16 bits.



Nuestra mención honorífica en esta ronda, Via Voice Pro Millenium Edition de IBM, ofrece una amplia variedad de funcionalidad tanto para aplicaciones de dictado como de control, además de las mejores calificaciones de precisión del grupo. Los novatos empezarán a trabajar con rapidez, mientras que los usuarios de poder apreciarán las capacidades avanzadas de macros del programa. No obstante, los evaluadores encontraron que algunas de las características de edición de Via Voice son complicadas, un factor determinantes en su facilidad de uso en general. La instalación y el manejo inicial transcurrieron sin problemas gracias a las pantallas del asistente bien diseñadas y un personaje animado que explica cómo empezar. Los usuarios podrán empezar a dictar de 20 a 30 minutos después de la instalación del software, lo que incluye de 10 a 15 minutos de lectura del texto del registro.



La calificación en capacitación refleja la instalación del programa, la facilidad y rapidez del inicio, la instalación del audio, así como los métodos y la extensión del aprendizaje de la voz. La calificación en dictado y corrección refleja los métodos de dictado, corrección con voz y formato con voz en las aplicaciones. La calificación en integración de las aplicaciones refleja la fuerza del producto para aceptar el dictado en aplicaciones de comandos y controles distintos al procesamiento de texto (como el escritorio de Windows, hojas de cálculo y visualizadores Web). La calificación en comandos y controles refleja la capacidad de la aplicación para utilizar comandos de voz a fin de navegar entre menús y cuadros de diálogo, así como su capacidad de crear macros robustas para comandos y controles.

	<b>Dragon Naturally-Speaking Preferred</b>	<b>FreeSpeech 2000</b>	<b>L&amp;H Voice Xpress Professional</b>	<b>ViaVoice Pro Millennium Edition</b>
Capacitación	Excelente	Aceptable	Buena	Excelente
Dictado y corrección	Excelente	Aceptable	Buena	Buena
Integración de las aplicaciones	Buena	Buena	Buena	Buena
Comandos y controles	Buenos	Buenos	Buenos	Excelente

En nuestras pruebas, Via Voice fue el producto más preciso: después del registro fue de 95%, que aumento a un sobresaliente 98% cuando dictamos el mismo documento después de hacer correcciones. En el uso diario, la mayoría de los usuarios deben tener una precisión entre estos porcentajes.

Via Voice, al igual que los otros productos que evaluamos, le permite controlar los menús y cuadros de diálogo en la mayor parte de los programas de Windows, con comandos en lenguaje natural disponibles para Microsoft Word y Excel, entre otros. Via Voice también le permite navegar el Web al activar Internet Explorer con la voz. Naturally Speaking también le permite decir los nombres de los vínculos, pero sólo Via Voice numera los elementos de la pagina, de modo que puede decir el nombre del elemento requerido.

Con todo, es complicado corregir los errores en Via Voice. El manual del programa sugiere emplear una combinación de mouse, teclado y voz para editar, en lugar de usar sólo comandos de voz.

Asimismo, es imposible emplear comandos de formato común en SpeakPad, el propio procesador de texto sencillo de Via Voice. Según las experiencias de los evaluadores, comandos como convertir en negritas las tres últimas palabras únicamente funcionaron en Microsoft Word.

A las excelentes calificaciones de precisión de Via Voice les afectan las limitadas capacidades de corrección del programa. Pero quienes son nuevos en el reconocimiento de

voz y los usuarios de poder que valoran las capacidades de macros del programa y su capacidad de guardar audio para reproducirlo después, encontrarán muchas características agradables en Via Voice.

## Reconocimiento del habla

Los sistemas comerciales han estado disponibles desde 1990. A pesar del aparente éxito de estas tecnologías, muy pocas personas utilizan el sistema del reconocimiento del habla en sus computadoras. Parece ser que muchos de los usuarios utilizan el ratón y el teclado para guardar o redactar documentos, porque les resulta más cómodo y rápido a pesar del hecho de que todos podemos hablar a más velocidad de la que tecleamos. Sin embargo, mediante el uso de ambos, el teclado y el reconocimiento del habla, nuestro trabajo será mucho más efectivo.

Este sistema donde está siendo más utilizado es en aplicaciones telefónicas: agencias de viajes, atención al cliente, información etc. La mejoría de estos sistemas de reconocimiento del habla han ido aumentando y su eficacia cada vez es mayor.

Los programas comerciales más famosos son:

a) Sistemas de dictado (dependientes del locutor):

- Via Voice de IBM- Nuance
- Dragon Naturally Speaking de Nuance
- Philips FreeSpeech
- Protile Live from NINSIGHT

b) Sistemas telefónicos (independientes del locutor):

- Nuance 8.5
- Telefónica : Software vocal de Telefónica
- Telisma

En software libre hay:

- CVoiceControl Se graba la orden como entrenamiento.
- PerlBox Sin entrenamiento, pero en inglés.
- Sphinx, del Sphinx Group en Carnegie Mellon University
- Open Mind Speech, antiguamente FreeSpeech