

# El programa de reconocimiento de voz Dragon Naturally Speaking

Lorenzo Serrahima\*

**Resumen:** Dragon Naturally Speaking es un programa de reconocimiento de voz que constituye una herramienta útil para los traductores porque permite dictar textos al ordenador en lugar de teclearlos. El entrenamiento previo es muy breve, y el aprendizaje del programa resulta muy intuitivo. Permite dictar en español o en inglés, indistintamente, con mucha precisión, revisar el texto dictado o dejar que el programa nos lo lea en voz alta, e incluso permite trabajar sin tener el ordenador delante. Sus inconvenientes principales residen en que exige una mayor atención a la hora de revisar los textos, solamente funciona con el sistema operativo Windows, tiene algunas incompatibilidades con los programas que trabajan con memorias de traducción y exige trabajar en soledad.

**Palabras clave:** reconocimiento de voz, programa Dragon Natural Speaking, ventajas e inconvenientes.

## A Voice-recognition Software: Dragon Naturally Speaking

**Abstract:** Dragon Naturally Speaking is a voice-recognition software which represents a highly useful tool for translators. The user can dictate the texts into the computer instead of typing them. The training period required is short and user-friendly. The program allows the user to dictate in either Spanish or English with a high level of accuracy, to edit the text or to request the system to perform a back-reading aloud. It even allows to work away from the computer. Its main drawbacks are the additional attention required for editing, the exclusive Windows compatibility, the incompatibility with some TM applications and the need to work in an isolated environment.

**Key words:** voice recognition, Dragon Natural Speaking software, advantages and drawbacks.

**Panace@ 2009; 10 (29): 72-75**

## 1. Introducción

Los programas de reconocimiento de voz son programas capaces de identificar las palabras dictadas a un micrófono y convertirlas en texto escrito o en órdenes para el ordenador. No son propiamente una herramienta de traducción, pues escriben exclusivamente en el mismo idioma en que se les dicta, pero permiten escribir todo tipo de textos sin necesidad de teclear. En la actualidad, existen en el mercado varias marcas, si bien las más conocidas son tres: Dragon Naturally Speaking (DNS), de la empresa Nuance, Via Voice (VV), de IBM, y Windows Vista. Todas ellas tienen en común que solo funcionan con el sistema operativo Windows. DNS ha ido mejorando con el tiempo y sigue manteniéndose como un programa útil de uso particular o para pequeñas empresas. VV también ha mejorado con diversas versiones, pero ha evolucionado hacia un programa con aplicaciones para grandes empresas y uso en red. Windows Vista ha salido recientemente al mercado con una herramienta de reconocimiento de voz con un nivel de calidad bastante prometedor (véase la comparativa entre DNS y el reconocimiento de voz de Windows Vista en las págs. 76-79 de este mismo número). En este artículo se abordan el programa DNS y sus características esenciales.

## 2. Características del programa DNS 9 preferred en español

El presente artículo se centra en la versión 9 de DNS en español, que funciona exclusivamente con el sistema operativo

Windows. En concreto, las versiones 9.10.000.141.SP1 (para Windows XP) y 9.51.250.254 (para Windows Vista). DNS permite dictar en inglés o en español, con diversas variantes: inglés de EE. UU., del Reino Unido, de Australia, de la India o del sudeste asiático, y español de España o de América Latina. El fabricante es Nuance Communications, Inc. (Estados Unidos) y en su página web<sup>1</sup> ya anuncia el lanzamiento de la versión 10, al igual que en la página española de la empresa.<sup>2</sup>

El programa trabaja fundamentalmente comparando tres bases de datos: la primera contiene muestras de voz del usuario que el programa ha asociado a distintas sílabas; la segunda contiene un vocabulario que el usuario puede ir ampliando a medida que va trabajando con el programa, y la tercera consiste en un corpus que incorpora numerosas expresiones que el propio usuario ha utilizado en documentos anteriores, así como una estadística del número de veces que ha dictado determinadas palabras. Durante el dictado, el programa recurre a la primera base de datos para identificar las correspondientes sílabas y, a continuación, a la segunda base de datos para identificar las palabras que corresponden a las diferentes secuencias de sílabas identificadas. A medida que las identifica, las escribe en la pantalla, si se trata de una palabra o un símbolo (p. ej., «coma»), o cumple una orden cuando se trata de esto último (p. ej., «nueva línea» introduce un salto de línea). En caso de duda entre dos palabras o expresiones fonéticamente similares (p.ej., *de la Seo* o *del aseo*), el programa recurre al corpus para comprobar con qué frecuencia

\* Veterinario, traductor autónomo. Barcelona (España). Dirección para correspondencia: [serrahima@gmail.com](mailto:serrahima@gmail.com).

se utiliza una u otra en el contexto que se está dictando y escoge la más probable. El programa puede escribir en su propio editor de texto (llamado Dragonpad) o prácticamente en cualquier otro programa del entorno Windows con el que habitualmente se pueda escribir con un teclado, ya sea un programa protegido por patente o uno de código abierto (MS Word, Excel, PowerPoint, Frontpage, Explorer, bloc de notas, OpenOffice...).

El programa se puede adquirir en la página web del fabricante de distintas formas: grabado en un CD, que se envía junto con unos auriculares y el clásico libro de instrucciones, o descargado directamente de la red.

Existen tres variantes: *standard*, *preferred* y *professional*. La más sencilla, *standard*, solamente permite dictar textos, números y órdenes en diversos programas. La versión intermedia, *preferred*, permite además que el programa reproduzca con voz propia lo que se le ha dictado previamente, es capaz de escribir un dictado a partir de una grabación en formato digital y también permite importar y exportar archivos de usuario. La versión más avanzada, *professional*, posibilita además trabajar en red, con macros, y disponer de varios vocabularios especializados por usuario. En realidad, esta última característica también la tiene la versión *preferred*, pues permite dar de alta un número ilimitado de usuarios, con lo que basta con crear un usuario para cada vocabulario especializado.

El precio varía sensiblemente de una versión a otra. Unos 100 € para la versión *standard*, de 150 a 300 € la versión *preferred* (según variantes) y unos 800 € la versión *professional*. La versión más adecuada para un traductor autónomo es la *preferred*. Por eso, a partir de aquí, solo se hará referencia a esta versión.

### 3. Instalación

Para poder instalar el programa se necesita un ordenador con sistema operativo Windows XP o Vista, con un procesador de 1 GHz por lo menos, 512 MB o 1 GB de RAM, 1 GB de espacio libre en el disco duro, una tarjeta de sonido con grabación de 16 bits, lector de CD (innecesario si se ha adquirido el programa directamente en la red) y conexión a Internet para activar el programa. Durante el proceso de instalación, el programa verifica que el ordenador cumpla con los requisitos mínimos; de no ser así, no se instala.

La caja del programa contiene un CD y un micrófono con auriculares. Podemos utilizar ese mismo micrófono o cualquier otro que nos interese, pero para obtener un buen rendimiento del programa resulta fundamental que el micrófono mantenga siempre la misma distancia con la boca. De ahí que no sea aconsejable utilizar un micrófono de mesa ni uno integrado en el propio ordenador. Tenido esto en cuenta, se conecta el micrófono al ordenador, se introduce el CD n.º 1 en el lector y se deja que el programa empiece con la instalación automática. Se selecciona la opción de instalación que se desee (típica o específica del usuario). En un momento dado el programa pide el CD n.º 2, y al final el programa nos pedirá que lo activemos. Esta activación consiste en enviar el número de serie (que nos han proporcionado con el programa) junto

con la huella dactilar de la máquina (que el propio programa busca e identifica) a la página web del fabricante. Con estos datos, la página genera una clave de activación que consiste en 60 caracteres (números y letras) agrupados en 15 subunidades de 4 caracteres cada una. La clave se introduce, mediante un simple proceso de copiar y pegar, en las casillas correspondientes del programa, con lo que este se activa definitivamente. El proceso puede hacerse de forma automática o manual, pero es recomendable hacerlo de forma manual, porque eso nos permite guardar una copia de la clave. Si se hace de forma automática, el programa también se activa, pero no nos queda constancia de la clave en ningún sitio, con el riesgo que eso implica si en el futuro tenemos que reinstalar el programa (algo que hay que tener muy presente en Windows). Cada programa comprado se puede instalar en cinco ordenadores distintos con sus correspondientes activaciones. Finalmente, el producto se registra en la página web para tener derecho a las futuras actualizaciones y a las comunicaciones comerciales de la empresa.

### 4. Cómo se empieza a utilizar el programa Entrenamiento

Una vez instalado el programa, lo primero que deberemos hacer será crear un usuario, pues lo pide el programa cuando se inicia por primera vez. El proceso es simple: se le da un nombre y se selecciona el idioma en que se va a dictar, con su correspondiente variante (p. ej., español de España). El propio programa ofrece por defecto una fuente de dictado, que suele ser el micrófono conectado al ordenador. A continuación se nos indica cómo colocar bien el micrófono para que se mantenga siempre a la misma distancia de la boca. El siguiente paso consiste en una prueba de volumen, en la que el programa comprueba que recibe la voz del usuario con el volumen suficiente. La prueba consiste en leer un pequeño párrafo que aparece en la pantalla y que, en condiciones normales, apenas dura 15 segundos. Después se realiza una prueba de calidad de sonido similar a la anterior. También consiste en leer un pequeño párrafo que aparece en la pantalla y dura otros 15 segundos, aproximadamente. Una vez comprobados el volumen y la calidad, el programa entra en la etapa de entrenamiento del usuario, que consiste en leer un texto algo más largo, que también ofrece el propio programa. El usuario tiene la posibilidad de escoger entre varias opciones, cualquiera de ellas de unos 10 minutos de duración. Es conveniente leer esta última prueba con la soltura habitual, porque eso aumentará después la precisión del programa: si se lee con mucho detenimiento, el programa cometerá muchos errores cuando se le dicte a la velocidad habitual. Todas estas pruebas son imprescindibles (sin ellas el programa no funciona), y una vez terminadas, ya se puede empezar a dictar.

Cuando se trabaja con él, el programa va escuchando la voz del usuario e inmediatamente va colocando en la pantalla las palabras que entiende. DNS incorpora una serie de órdenes propias que facilitar al dictado: algunas se refieren a los signos de puntuación (por ejemplo, al dictar la palabra «coma» en pantalla aparece el símbolo «,») y otras son órde-

nes de edición. Esto acarrea la segunda parte del aprendizaje necesario de este programa: las órdenes.

Afortunadamente, el aprendizaje de las órdenes resulta breve, puesto que todas ellas son muy intuitivas. Por ejemplo, para hacer que el cursor pase a la línea siguiente, la orden es «nueva línea», y para que la palabra siguiente aparezca escrita con mayúscula, la orden es «mayúscula inicial». También es factible controlar con la voz todas las posibles funciones del programa de tratamiento de texto, pero suele resultar mucho más cómodo controlar con la voz las funciones más habituales y seguir realizando las menos usuales con el teclado, el ratón o una combinación de ambos. Una de las características fundamentales y más útiles de este programa consiste en que su uso no excluye la funcionalidad del teclado, es decir, que siempre se puede teclear cualquier cosa mientras se está utilizando el programa de dictado. En determinadas ocasiones será necesario que en la pantalla aparezca la palabra dictada y no el símbolo ortográfico. En estos casos, lo más rápido y lo más sencillo consiste en corregir la palabra directamente con el teclado.

### **Los errores no se detectan automáticamente**

Al empezar a utilizar el programa se puede alcanzar una precisión de entre el 80 y el 90 % de las palabras, según la complejidad del texto. Es decir, se identifican correctamente entre el 80 y el 90 % de las palabras y órdenes dictadas con el entrenamiento mínimo. El inconveniente radica en que el programa siempre escoge una palabra correctamente escrita, aunque esté totalmente fuera de contexto: no piensa ni entiende nada de lo que está escribiendo (no busca la contextualización de la palabra) y, por lo tanto, puede escribir auténticas barbaridades sin ningún problema. De modo que si uno dicta la palabra *camión* y el programa lo entiende mal, puede ser que escriba la palabra *acción*, pero no escribirá nunca una palabra con una ortografía incorrecta, digamos *kamón*. Esto obliga a prestar mucha atención durante la revisión posterior del texto, puesto que el corrector ortográfico habitual en muchos programas de tratamiento de textos no identificará la palabra como errónea, porque ciertamente estará bien escrita. El usuario deberá recordar en todo momento que el programa no piensa, así que no tiene ninguna referencia para distinguir la expresión «Juan piensa *en volver* la caja del revés» de la expresión «Juan piensa *envolver* la caja del revés». Es precisamente a este tipo de palabras o expresiones homófonas a las que más atención se deberá prestar durante la revisión del texto. Determinadas palabras tienen tendencia a aparecer de forma errónea con más facilidad que otras, por ejemplo las siguientes:

- El ↔ él
- Por qué ↔ porque
- Éste ↔ este ↔ esté
- Hay ↔ ahí ↔ ¡Ay!
- Valla ↔ vaya

Evidentemente, no se trata de una lista exhaustiva, y cada usuario deberá identificar qué palabras le aparecen de forma errónea con mayor frecuencia para crear sus propios sistemas

de control que las detecten y corrijan en el texto final. En cuanto a palabras que aparecen de forma incorrecta por ser desconocidas para el programa, existe una orden muy simple que permite corregirlas y, además, incorporar la palabra corregida al vocabulario específico del usuario. De esta manera, el vocabulario se va ampliando y el índice de errores disminuye.

También existe la posibilidad de que el propio programa «lea» los archivos de texto que el usuario tenga escritos y archivados en su ordenador, identifique todas las palabras que contienen y que no estén incluidas en el propio vocabulario del programa, haga una lista con ellas y ofrezca al usuario la posibilidad de incorporarlas a su vocabulario. De esa forma, el índice de precisión puede llegar a superar el 95 %. Gracias a esta utilidad, se puede incorporar cualquier palabra que haya utilizado con anterioridad, por compleja que sea. A modo de ejemplo, las siguientes palabras se han dictado con el programa DNS:

bacterias fotoautótrofas, metanógenas, termoacidófilas, reductoras de sulfatos, esporóforos bacterianos, bacterias procariotas y eucariotas, beocitos, plástidos fotosintetizadores.

En realidad, cuanto más compleja sea la palabra incorporada, más difícil será que el programa se equivoque después al dictarla, puesto que difícilmente existirá otra que se pronuncie de forma tan parecida como para inducir a error.

### **Ergonomía y compatibilidad del DNS**

La mayor utilidad para el traductor profesional radica en que, al dictar los textos en lugar tenerlos que teclear en el ordenador, se pueden adoptar posturas mucho más cómodas y relajadas que aquellas a las que obliga el teclado. Además, se puede ir cambiando de postura varias veces durante el día, lo que reduce el cansancio y evita lesiones a largo plazo. Por otro lado, ir hablando tranquilamente cansa bastante menos que teclear, lo que permite aumentar sensiblemente la productividad diaria. Existe una variante del programa que admite micrófonos inalámbricos con la tecnología Bluetooth (y que cuesta unos 100 € más). Ello permite al usuario incluso levantarse y pasear por la habitación mientras dicta. Este autor no tiene experiencia con esos micrófonos, y por lo tanto no sé si mantendrán el mismo grado de precisión o no. Por otro lado, un micrófono con un cable de 2 m proporciona una gran movilidad a un coste muy inferior.

Es cierto que después habrá que dedicar más tiempo a revisar el trabajo realizado, pero, en conjunto, el volumen de trabajo que se puede hacer en un día puede ser más del doble del que se hace con el teclado. Una de las características que pueden limitar la utilidad del programa es el hecho de estar hablando continuamente: eso exige trabajar solo en una habitación. Sin embargo el programa es bastante insensible a los ruidos ambientales y a la música: se puede trabajar perfectamente escuchando música sin que el programa pierda ni un ápice de precisión.

En cuanto a su compatibilidad con otros programas habituales entre los traductores, como los basados en memorias de traducción, no está del todo bien resuelta. Presenta incompati-

bilidades bastante acentuadas con Wordfast y alguna incompatibilidad menor con Trados. Sin embargo, es un programa ideal para la traducción de textos largos y poco repetitivos, como libros o artículos.

### Revisión final

Una vez se ha acabado de dictar todo el texto que se desea, existen dos formas de revisarlo: releer todo a la manera habitual o dejar que el propio programa lea con su propia voz aquello que está escrito en pantalla. Es una voz femenina muy humana, que en nada nos recuerda que lo está leyendo una máquina. Además, se puede ajustar la velocidad de lectura al gusto del usuario. La ventaja de esta utilidad es que permite descansar la vista; el inconveniente es que, para que la lectura sea comprensible, ha de ir relativamente lenta. Esta utilidad también se puede aplicar a cualquier otro texto editado con un editor de textos, tanto si ha sido dictado por el usuario como si no.

### 5. Dictado sin ordenador

Cuando el usuario haya aprendido las órdenes más habituales de dictado, podrá aprovechar otra de las funciones del programa: el dictado sin ordenador. El programa tiene una función de reconocimiento de grabaciones que permite que el usuario vaya dictando el texto a algún dispositivo digital de grabación. Eso permite trabajar fuera del lugar habitual de trabajo, incluso mientras se viaja en transporte público. Después bastará con descargar la grabación en el ordenador

y dejar que el programa la transcriba. DNS es capaz de reconocer grabaciones en formato WAV, si bien se debe ajustar la velocidad a 21 kbps. El índice de precisión de este tipo de grabaciones difícilmente llega a los mismos valores que se alcanzan con el micrófono de los auriculares, puesto que, por poco que el usuario se mueva, la distancia desde su boca hasta la grabadora varía ligeramente, lo que va en detrimento de la precisión. Por lo tanto, los textos así dictados requerirán algo más de tiempo de revisión, pero siempre permitirán aprovechar un tiempo que el usuario no ha pasado delante de su ordenador. Para las personas que tienen que hacer desplazamientos periódicos en transporte público, sobre todo en las grandes ciudades, esta es una característica nada desdeñable.

### 6. Conclusión

En resumen, el programa DNS *preferred* en español es una herramienta útil para un traductor, porque permite dictar textos al ordenador, tanto en español como en inglés, con mucho menor esfuerzo que a través del teclado, y además permite trabajar sin ordenador. Sin embargo, tiene el inconveniente de que solamente funciona en entorno Windows y exige mucha más atención a la hora de revisar los textos dictados. En conjunto es una opción muy recomendable.

### Notas

1. <[www.nuance.com/products/](http://www.nuance.com/products/)>.
2. <<http://spain.nuance.com/>>.

## Inclusa

José Ignacio de Arana

Médico y escritor. Madrid (España)

Los lugares para acoger a niños abandonados, muchas veces al poco de nacer, o huérfanos sin cobijo ni otra ayuda, condenados por tanto a morir, fueron desde su origen instituciones creadas por la Iglesia católica y por algunas personas directamente influidas por su pertenencia a ella. Surgieron pronto tres tipos: *xenodoquios* o casas de amparo y hospicios de maternidad para mujeres pobres en trance de dar a luz; *derephotrofias* para los recién nacidos abandonados, y *horfanotrophias* o casas para huérfanos desamparados. Los principales hospicios de Occidente surgen, sin embargo, en la Edad Media. Casi todos ellos son solo dependencias de hospitales —en el sentido medieval, tan restrictivo, del término— atendidos por mujeres con mejor voluntad que ciencia y solo en ocasiones con algún médico.

Hacia 1567 se crea en Madrid, en el convento de la Victoria, situado junto a la Puerta del Sol, la cofradía de Nuestra Señora de la Soledad y las Angustias con ese fin caritativo. Pocos años después, en el lejano Flandes iba a tener lugar un episodio sin aparente relación con lo anterior. Al conquistar los tercios españoles la ciudad flamenca de Enkuisen, encontraron que los enemigos protestantes habían profanado todas las iglesias; un soldado halló en una de ellas un cuadro de la Virgen de la Paz rodeada de ángeles y con un niño a sus pies, y decidió guardarlo en su equipaje militar. Tras su regreso a España, en una de las audiencias que don Felipe II concedía a los veteranos, aquel soldado regaló al rey la imagen rescatada, y el monarca, viendo la escena del niño a los pies de la Virgen, decidió a su vez donarla a la cofradía que en la capital se dedicaba a cuidar niños abandonados. En el convento de la Victoria fue entronizada y pronto fue objeto de una enorme devoción entre los madrileños. Pero estos no sabían pronunciar el nombre de aquella ciudad y comenzaron a utilizar para el cuadro la advocación de «Virgen de la Inclusa». Poco a poco, esta palabra sustituyó en el habla popular al nombre del convento y cofradía, pasando esta a denominarse simplemente *inclusa*. La nueva denominación hizo fortuna y de allí se extendió a todas las instituciones españolas dedicadas como ella a la recogida de expósitos. El nombre de *hospicio* se reservó para los centros que se hacían cargo de niños mayorcitos y también de los que, habiendo comenzado su vida en la inclusa, lograban sobrevivir más allá de cierta edad.

Reproducido con autorización de «El laboratorio del lenguaje», de *Diario Médico*. <[www.diariomedico.com/blogs/laboratorio](http://www.diariomedico.com/blogs/laboratorio)>