NUEVAS PROPUESTAS TECNOLÓGICAS PARA LA PRÁCTICA Y EVALUACIÓN DE LA PRONUNCIACIÓN DEL ESPAÑOL COMO LENGUA EXTRANJERA.

DAVID ESCUDERO

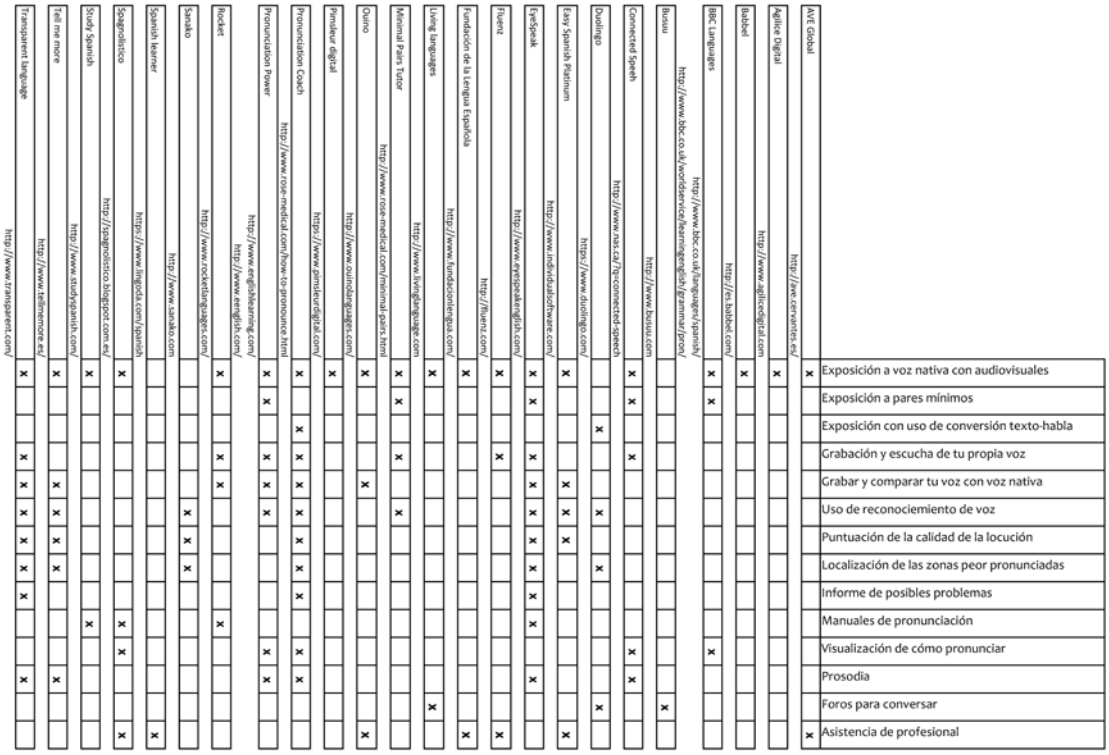
Departamento de Informática, Universidad de Valladolid

MARIO CARRANZA

Departamento de Filología Española, Universidad Autónoma de Barcelona

RESUMEN

La pronunciación, a pesar de ser uno de los aspectos más importantes para el dominio de una lengua extranjera, no ha sido estudiada con la misma profundidad que otros aspectos lingüísticos como la gramática o el léxico, lo que se refleja en el hecho de que en los manuales de enseñanza raramente aparecen ejercicios de pronunciación. Por otro lado, los avances tecnológicos permiten incorporar las tecnologías de la información en el aprendizaje de lenguas extranjeras. En este artículo se describe el estado de la cuestión de la enseñanza de la pronunciación asistida por ordenador a partir de una revisión de los modelos de actividades incluidos en un conjunto de aplicaciones y programas seleccionados, cuyo fin específico es la mejora de la pronunciación: uso de voz grabada para la mejorar la percepción de los sonidos en lengua extranjera; uso de habla sintetizada para sustituir las grabaciones; grabación y escucha de la propia voz (del estudiante), bien sea de forma aislada, combinada con voz nativa o manipulada para parecer más nativa; uso de reconocimiento de habla con palabras o frases previamente establecidas; uso de reconocimiento de habla con dominio abierto, incluido en un sistema de diálogo o en un sistema de dictado. La capacidad de adaptarse a los cambios tecnológicos, de definir metodologías adecuadas y de demostrar su utilidad y eficacia, son las claves para el éxito futuro de estas herramientas.



	AVE Global	Agilice Digital	Babbel	BBC Languages	Busuu	Connected Speech	Duolingo	Easy Spanish Platinum	EyeSpeak	Fluentz	Fundación de la Lengua Española	Living languages	Minimal Pairs Tutor	Ouino	Pimsleur digital	Pronunciation Coach	Pronunciation Power	Rocket	Sonako	Spanish learner	Soagnolistico	Study Spanish	Tell me more	Transparent language
	http://ave.cervantes.es/	http://www.agilicedigital.com	http://es.babbel.com/	http://www.bbc.co.uk/languages/spanish/ http://www.bbc.co.uk/worldservice/learningenglish/grammar/pron/	http://www.busuu.com	http://www.rosa.ca/fp=connected.speech	https://www.duolingo.com/		http://www.individualsoftware.com/	http://www.pretpeakenglish.com/	http://fluentz.com/	http://www.fundacionlengua.com/	http://www.livinglanguage.com	http://www.rose-medical.com/minimal-pairs.html	http://www.quinolanguages.com/	https://www.pimsleurdigital.com/	http://www.rose-medical.com/how-to-pronounce.html	http://www.englishlearning.com/ http://www.eenglish.com/	http://www.rocketlanguages.com/	http://www.sanako.com	https://www.linguoda.com/spanish	http://spagnolistico.blogspot.com.es/	http://www.studyspanish.com/	http://www.tellmemore.es/	http://www.transparent.com/
×	Exposición a voz nativa con audiovisuales	×	×	× ×	×	×		×	×	×	×	×	×	×	×	×	×	×			×	×	×	×
	Exposición a pares mínimos			×		×	×		×				×	×		×	×								
	Exposición con uso de conversión texto-habla						×									×	×								
	Grabación y escucha de tu propia voz					×		×	×	×	×		×	×		×	×	×						×	
	Grabar y comparar tu voz con voz nativa							×	×		×			×		×	×	×						×	
	Uso de reconocimiento de voz						×		×		×		×			×	×		×					×	
	Puntuación de la calidad de la locución							×	×		×					×	×		×					×	
	Localización de las zonas peor pronunciadas						×		×		×					×	×		×					×	
	Informe de posibles problemas								×		×					×	×							×	
	Manuales de pronunciación								×								×					×			
	Visualización de cómo pronunciar			×		×			×								×							×	
	Prosodia						×																		
	Foros para conversar				×				×																
×	Asistencia de profesional										×			×											

Tabla 1: Tipos de actividades presentes en los programas analizados.

1. INTRODUCCIÓN

Bernstein (2003) argumenta que la pronunciación posee una importancia fundamental para la comunicación; pero, a pesar de esto, no ha sido valorada adecuadamente en los cursos de aprendizaje de lenguas extranjeras. Los productos TIC son, a priori, una herramienta que puede ser de gran utilidad en la práctica de la pronunciación porque permiten definir ejercicios que potencian la exposición a la lengua no nativa así como su producción. La revisión que presentamos aquí, se centra en extraer el conjunto de actividades típicas para practicar la pronunciación con estos programas.

Aunque presentamos una lista nominal de programas, este artículo no pretende juzgarlos ni hacer un ranking. Por el contrario, nuestra aproximación es puramente analítica y descriptiva para finalizar discutiendo sobre los factores que posiblemente hagan que los programas tengan éxito o desaparezcan a corto o medio plazo. Primero, listamos las herramientas analizadas detallando los tipos de actividades típicas que incluyen para la

práctica de la pronunciación. A continuación, discutimos los aspectos que estos productos deberán tener en cuenta para tener éxito y mantenerse en el mercado.

2. PRÁCTICA DE LA PRONUNCIACIÓN ASISTIDA POR ORDENADOR

Se han analizado 23 programas que ofrecen cursos de español como lengua extranjera y que incluyen actividades relacionadas con la pronunciación (ver tabla 1). A continuación, detallamos a qué hace referencia cada tipo de actividad concreta distinguiendo entre actividades de exposición, de discriminación perceptiva, de producción, de instrucción y actividades sociales.

2.1. Actividades de exposición

En las actividades de exposición, el alumno escucha diferentes enunciados en la lengua meta. Distinguimos entre las actividades que utilizan material audiovisual pregrabado para la exposición, y aquellas que emplean voz sintetizada. En las actividades de exposición al habla nativa empleando materiales audiovisuales es posible escuchar grabaciones de voz de hablantes nativos que el usuario puede reproducir cuando precise. Aunque es el más elemental de los tipos de actividad, no debe despreciarse, ya que en los últimos años se han incrementados los medios que permiten la exposición al habla nativa: canal dual en TV, películas en DVD, YouTube... Podemos pensar que el ordenador puede aportar poco, pero debemos tener en cuenta que la organización adecuada de todo el material puede ser clave para facilitar el acceso al mismo. Destacamos aquí el trabajo de *Agilice Digital*, cuyo producto estrella es un paquete de vídeos indexados para facilitar el acceso a los mismos desde las diferentes lecciones en función del contenido que se esté trabajando.

También podemos usar voz sintetizada para realizar la exposición a la voz nativa. La ventaja de utilizar un sintetizador de voz está en el hecho de que no es necesario tener

grabaciones realizadas previamente. El sintetizador de voz libera al desarrollador de depender de locutores para hacer las lecturas. A pesar de que la calidad de la voz sintética ha mejorado considerablemente durante la última década, son pocos los programas que emplean esta opción. De hecho, existen trabajos de investigación que son críticos al respecto (Handley, 2009) y que han podido influir en la escasez de productos que incluyen esta opción, a pesar de que la tecnología actual de la síntesis de habla ha alcanzado un nivel de calidad difícil de distinguir de la voz humana.

Destacamos no obstante la opción de *Duolingo* por emplear esta opción y el rápido crecimiento en el número de usuarios de la herramienta en los últimos años. La razón para utilizar voz sintetizada es la flexibilidad que permite poder producir ejercicios que se adaptan a las necesidades de cada usuario sin tener que depender de las restricciones de disponer o no de la voz grabada. La siguiente razón es el precio, ya que disponer de una máquina que pueda hacer las locuciones es mucho más económico a largo plazo que contratar a un locutor profesional para grabar los audios.

2.2. Actividades de discriminación perceptiva

Una actividad de discriminación consiste en someter al usuario a la exposición de un par de palabras que se diferencian en un solo fonema. El alumno aprende con estos ejercicios que es capaz de percibir los contrastes de fonemas de la lengua extranjera (Cámara Arenas, 2012). Hemos encontrado ejemplos de discriminación por pares mínimos para el inglés como *Oxford Pronunciation Files* , o *Minimal Pairs Tutor*. Esta última aplicación está preparada para el español pero no se ofrece en versión en línea. La aplicación de desarrollo propio que describimos en la discusión es un ejemplo de sistema que ofrece la posibilidad de discriminación de pares mínimos para el español (Escudero *et alii*, 2015).

2.3. Actividades de producción

En las actividades de producción se hace hablar al usuario. Distinguimos aquellas que se limitan a grabar al usuario, permitiéndole que compare su propia voz con la voz de un locutor nativo, de aquellas actividades que incluyen técnicas de procesamiento del habla, entre otras, técnicas de reconocimiento automático del habla (RAH).

Con la grabación y escucha de su propia voz, el usuario puede escucharse a sí mismo y compararse con una locución de referencia realizada por un locutor nativo. El ordenador, frente al uso de grabadoras convencionales, aporta flexibilidad, permitiendo acceder a ejemplos con los que compararse de manera sencilla. Los ejercicios de grabación de la propia voz y repetición no sirven de mucho si el sistema no indica cuál es la diferencia entre el modelo y la grabación del estudiante o dónde se encuentran los errores. Esta “concepción” de la pronunciación presupone que el estudiante será capaz por sí mismo de decidir el grado de semejanza de su propia voz con el modelo, lo que, en la mayoría de los casos, no sucede, puesto que muchos contrastes entre fonemas son imposibles de diferenciar por hablantes no nativos (por ejemplo, el caso de /ɾ/- /l/ por estudiantes japoneses). La teoría de la “sordera fonológica", introducida en los años 30 por Trubetzkoy (1939), parece que no ha tenido la repercusión necesaria para cambiar totalmente el modelo (Llisterri 2001, 20-1).

Nos llama la atención que algunas herramientas ofrecen la posibilidad de hacer *shadowing*, reproduciendo simultáneamente su voz y la voz nativa de referencia. Este es el caso de la herramienta *Rocket*. El objetivo es que el estudiante repita lo que escucha intentando imitar lo máximo posible la voz del locutor nativo.

Proliferan las herramientas que utilizan RAH, habiendo analizado alrededor de una decena de sistemas que incluyen esta funcionalidad. El uso es muy variado

distinguiendo primero herramientas que se limitan a aplicar reconocimiento automático del habla sobre las producciones de voz del alumno para comprobar que este da una respuesta correcta (como *Rosetta Stone,* o *Duolingo).* Los sistemas de reconocimiento de habla son el recurso estrella en los programas de pronunciación asistida por ordenador (Dalby, y Kewley-Port, 2013). El uso de un vocabulario cerrado y de unas frases previamente establecidas permite contrastar la salida obtenida por el reconocedor de habla con la salida esperada. Se trata de un resultado obtenido automáticamente que puede ser utilizado para emitir juicios sobre la calidad de la voz. Además, los sistemas de reconocimiento de habla operan sobre la base del cómputo de probabilidades y puntuaciones que también pueden ser utilizadas para estimar la calidad de la voz, especialmente si se han modelado las variantes de pronunciación en el diccionario.

Para incrementar la motivación del usuario, encontramos también soluciones que simulan un diálogo entre el usuario y agentes simulados (por ejemplo *Easy Spanish).* El número de posibles respuestas permitidas al usuario está limitado, no siendo pues sistemas de diálogos que permiten una flexibilidad razonable que incremente la riqueza del diálogo. En un sistema de diálogo se establecen turnos entre la máquina y la persona. El motivo del diálogo o pseudo-diálogo suele ser la resolución de una tarea. La máquina interpreta la locución del usuario para generar nuevos turnos y continuar la conversación. Estos sistemas evitan la última de las limitaciones mencionada en el párrafo anterior. Si se flexibiliza el rango de frases que puede decir el usuario, se mejora con ello también el número de actividades que pueden ser diseñadas. Si el diálogo se desarrolla sobre un asunto de interés para el alumno (por ejemplo una entrevista de trabajo), puede crearse un entorno comunicativamente significativo que aumente las opciones de éxito porque aumenta la motivación y el interés de los usuarios.

Otros sistemas valoran la adecuación de la pronunciación puntuando la intervención del usuario. La figura 1 muestra el sistema *Easy Spanish* que utiliza un contador de velocidad simulado para juzgar la adecuación de la intervención del usuario y el evaluador de calidad de *Pronunciation coach.*

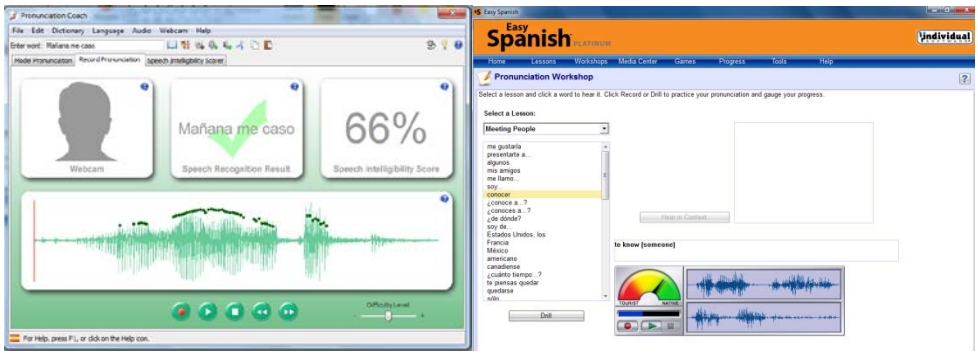


Figura 1: Interfaz con marcador de calidad de la locución.

Para evaluar automáticamente la calidad de la pronunciación no nativa se han definido índices como el GOP (Goodness Of Pronunciation) que computan la desviación de la pronunciación del estudiante con respecto a los modelos fonéticos de hablantes nativos (Kanters, 2009). Sin embargo el índice GOP presenta limitaciones para hacer un diagnóstico eficiente que permita informar al locutor sobre los problemas particulares que tienen sus locuciones (van Doremalen, Cucchiarini, y Strik, 2013). Como apunta Bernstein, antiguo director técnico del área CAPT (Computer-Assisted Pronunciation Training) de *Pearson Education*, el paso de sistemas de evaluación de pronunciación a sistemas de diagnóstico que puedan ayudar a los usuarios a mejorar sus resultados es el gran desafío actual de esta tecnología (Cheng, 2014)

Más interesantes resultan los sistemas que localizan las partes de la locución con peor pronunciación mediante alguna marca visual. Estos sistemas permiten encontrar las zonas que son, a juicio de la máquina, las peor pronunciadas e invitan a escuchar los fonemas marcados como incorrectos, pronunciados correctamente por un locutor nativo.

Las herramientas *EyeSpeak* y *Easy Speak* son buenos ejemplos que emplean estas técnicas. La herramienta EyeSpeak permite introducir un texto que después el usuario lee. El sistema nos informa de los fonemas peor pronunciados.

2.4. La prosodia

Algunos de los sistemas que incluyen análisis automático del habla, también se ocupan de la prosodia ayudando a localizar fragmentos mal pronunciados. Los sistemas *Tell me more* y *Pronunciation coach* incluyen la curva de entonación del locutor nativo y del usuario. El sistema *EyeSpeak* incluye intensidad y F0 entre los parámetros de valoración. Nos parece interesante este enfoque para el caso particular de la prosodia, porque, a diferencia de lo que ocurre con la parte segmental, aquí se aporta una retroalimentación visual que puede guiar al usuario en futuras repeticiones. Ante un perfil de F0 de un nativo y de un no nativo cuyo contorno prosódico difiere significativamente en la posición de determinados picos o valles es relativamente sencillo reaccionar y corregir la locución, aunque siempre será problemático decidir qué picos y valles son los que marcan realmente la entonación de ese enunciado, independientemente de las características prosódicas de cada hablante. Además, las ayudas visuales pueden resultar difíciles de interpretar si el usuario no está entrenado.

2.5. Instrucción explícita de la articulación de los sonidos

Por instrucción entendemos la especificación de un conjunto de pautas o consejos para articular correctamente el tracto vocal y la posición de los labios y poder así emitir de manera apropiada los sonidos. Distinguimos los manuales clásicos, y las animaciones multimedia. El manual más afamado en lo referente a la pronunciación del español es el redactado a mediados del siglo pasado por Navarro Tomas (1918). *AVE Global* o *Rocket,* entre otros, se apoyan en la especificación de este tipo de reglas para indicar la forma correcta de pronunciar. Los símbolos fonéticos se relacionan con imágenes y

explicaciones que indican cómo pronunciarlos. Las animaciones permiten representar el movimiento de los órganos articulatorios para pronunciar los sonidos de la lengua meta. Encontramos vídeos como los de la BBC (en inglés) y animaciones 2D en *Pronunciation Pro* o *Pronunciation Power*.

2.6. Aprendizaje social

Empleamos el término *social* para referirnos a aquellas herramientas que facilitan el contacto con otras personas, sean otros estudiantes o instructores. Redes sociales como *Busuu*, entre otras, permiten poner en contacto entre sí a personas que están aprendiendo una lengua extranjera. Esto permite incrementar la motivación del estudiante al poder, por ejemplo, participar en juegos serios y competir con otros usuarios. Destacamos la filosofía de *Duolingo*, que lleva esta idea al extremo de hacer que sean los propios usuarios los que creen sus propios cursos. Más práctica es la aproximación de sistemas como *Lingoda* o *LiveMocha* cuyo nicho de mercado es la utilización de nuevas tecnologías para poner en contacto a los usuarios con profesores especializados.

3. DISCUSIÓN

Las herramientas CAPT que hemos listado forman parte de un ecosistema en el que también participan profesores de lengua extranjera, alumnos y desarrolladores de plataformas tecnológicas. Como en todo ecosistema, la supervivencia es la principal preocupación de sus integrantes. Las distintas herramientas para la práctica de pronunciación compiten entre sí para atraer alumnos. Para ello, deben demostrar que son eficaces y, además, deben adaptarse a los constantes cambios de las plataformas. Como consecuencia, si las herramientas no se adaptan y no proponen nuevos usos pedagógicos de la tecnología existente, posiblemente, desaparecerán. Es necesario que los programas se adapten a la evolución de las plataformas que son (1) sistemas de

escritorio (2) navegadores web y (3) *apps* para dispositivos móviles. La elección de una u otra plataforma supone un alto número de condicionantes a los que los fabricantes deben adaptarse. Quizá el más crítico es el acceso al micrófono, un recurso controlado por el sistema operativo, cuyo control requiere tener en cuenta aspectos de privacidad de la información. Los diferentes sistemas operativos y los navegadores web van a tener mucho que ver en el éxito o fracaso de estos programas.

La segunda cuestión es la integración de la herramienta dentro de una metodología de aprendizaje del idioma. Muchos de los sistemas descritos separan la enseñanza de la pronunciación con herramientas CAPT del resto de unidades en lo que algunas llaman *Taller de pronunciación.* Otras, dan al reconocimiento de habla una importancia capital al aparecer en todas las lecciones como una actividad obligatoria (ej. *Rosetta Stone).* La definición de una metodología que se adapte a un perfil de usuario típico parece estar siendo la clave del éxito de alguno de estos sistemas, como *Duolingo.*

Por último, una clave importante para la supervivencia de los sistemas es que demuestren ser útiles. Para discutir sobre este particular presentamos los resultados que obtuvimos en un test de usabilidad de un producto de desarrollo propio (Escudero et alii, 2015). Se trata de un juego serio que invita a escuchar y producir pares mínimos empleando un sintetizador de voz y un sistema de reconocimiento automático de habla. Se realizaron pruebas con tres tipos de usuarios: nativos, estudiantes de filología y estudiantes de informática y se demostró que el sistema permite valorar la calidad del locutor en términos de su pronunciación al contrastar si el sistema de RAH identifica correctamente la palabra que debe pronunciar. Asimismo, se comprobó que el resultado que devuelve el RAH como indicador de calidad efectivamente es sintomático de la calidad del habla. No obstante, y a pesar de estos prometedores resultados, también se

puso en evidencia que la correcta elección de actividades es capital para no provocar efectos adversos en la sensación de utilidad del sistema.

4. CONCLUSIONES

En este artículo se ha realizado una revisión de aquellos programas que ofrecen herramientas para la mejora de la pronunciación. Como resultado, hemos identificado una serie de actividades tipo que han sido descritas. Es de esperar que en pocos años muchos de los programas analizados hayan desaparecido, apuntando como causas la necesidad de apartarse a unos soportes en constante cambio y la necesidad de integrarse en metodologías en las que demuestren su utilidad. La tecnología no debe imponerse como un fin, sino que es la metodología la que debe indicar cómo sacar el máximo provecho de las tecnologías lingüísticas para crear actividades motivadoras y adecuadas pedagógicamente al aprendizaje.

BIBLIOGRAFÍA

Bernstein, J. “Objective measurement of intelligibility”. *Proceedings of the ICPhS*. 1581-84, 2003.

Cámara Arenas, E. *Curso de pronunciación de la lengua inglesa para hispano-hablantes, a native cardinality method*. Valladolid: Secretariado de Publicaciones de la Universidad de Valladolid, 2012.

Cheng , J., Y. Z. D’Antilio, X. Chen y J. Bernstein. “Automatic Assessment of the Speech of Young English Learners”. Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Applications , 12, 2014.

Dalby, J., y D. Kewley-Port. “Explicit pronunciation training using automatic speech recognition technology”. *Calico Journal*, *16*(3), 425-45, 2013.

Escudero-Mancebo, D., E. Cámara-Arenas, E., Tejedor-García, C., González-Ferreras, y V. Cardeñoso-Payo. “Implementation and Test of a Serious Game Based on Minimal Pairs for Pronunciation Training”. *Proceedings of SLaTE 2015*, en prensa.

Handley, Z. “Is text-to-speech synthesis ready for use in computer-assisted language learning?” *Speech Communication*, 51(10), 906-19, 2009.

Kanters, S., C. Cucchiarini, y H. Strik. “The goodness of pronunciation algorithm: a detailed performance study”. *Proceedings of SLaTE 2009*,. 49-52, 2009.

Llisterri, J. “Enseñanza de la pronunciación, corrección fonética y nuevas tecnologías. *Es Espasa, Revista de Profesores*, 28 de noviembre de 2001.

Navarro Tomás, T. *Manual de pronunciación española* (22ª ed., 1985). Madrid: Consejo Superior de Investigaciones Científicas CSIC, 1918..

Seneff, S., C. Wang, y J. Zhang.. “Spoken conversational interaction for language learning”. *InSTIL/ICALL Symposium*, 2004.

Trubetzkoy, N. S. *Grundzüge der phonologie. Travaux du Cercle Linguistique de Prague,7*. Göttingen: Vandehoeck & Ruprecht, 1939.

van Doremalen, J., C. Cucchiarini, y H. Strik. “Automatic pronunciation error detection in non-native speech: The case of vowel errors in Dutch”. *The Journal of the Acoustical Society of America*, 134(2), 1336-47, 2013.