

Influencia del formato de presentación de la información lingüística en el aprendizaje trans-situacional de términos verbales nuevos

Pamela Lopes da Cunha, Silvano Zanutto y Alejandro Wainseboim

Introducción

Mientras que la adquisición del lenguaje en infantes ocurre por exposición al contexto lingüístico, en adultos se requieren procesos de aprendizaje conscientes y explícitos. Las bases neuropsicológicas de esta diferencia no se conocen con exactitud. En infantes, el proceso parece involucrar el análisis estadístico de la señal lingüística *per se* (Saffran, Aslin y Newport, 1996) y de la coocurrencia entre contexto visual y auditivo. Tanto en adultos como en infantes, este tipo de aprendizaje inferencial se ha denominado aprendizaje estadístico trans-situacional, y se ha propuesto como uno de los fundamentos del aprendizaje de términos nuevos durante la adquisición del lenguaje materno en la infancia (Siskind, 1996; Lacerda y Sundberg, 2006; Yu y Smith, 2007).

Dado que la adquisición de la lengua materna se da en un contexto ambiental con una estrecha relación entre interacción social y estímulo lingüístico (Lacerda y Sundberg, 2006; Yu y Ballard, 2007), este proceso parecería involucrar una asociación multimodal entre las distintas entradas sensoriales que procesa el individuo. En infantes de 4 a 8

meses de edad se ha visto que la capacidad de asociar estímulos auditivos novedosos con imágenes que representen acciones se encuentra significativamente por debajo de la capacidad de asociar formas lingüísticas nuevas con objetos (“sustantivos”) o atributos sensoriales como forma o color (“adjetivos”) (Marklund y Lacerda, 2006). Estos resultados indicarían una mayor dificultad implicada en el proceso de adquisición de verbos, comparado con sustantivos y adjetivos, coincidiendo con el análisis del orden de adquisición de los diferentes tipos de palabras en el lenguaje materno, donde la adquisición de sustantivos precede a la de verbos como clase gramatical (Gentner, 1982; D’Odorico y Fasolo, 2007).

En adultos, se ha visto que es posible adquirir significados verbales nuevos por medio de la presentación simultánea de acciones (movimientos) y formas ortográficas y fonológicas nuevas (Fernández Feijóo *et al.*, 2008).

En el presente capítulo presentamos la continuación del trabajo de Fernández Feijóo *et al.* (2008), planteándonos: a) estudiar si el formato de presentación de la señal lingüística modificará el grado de aprendizaje logrado; b) estudiar la neurofisiología del procesamiento de frases Correctas e Incorrectas mediante el registro de electroencefalografías (EEG) durante la fase de testeo. En particular, nos preguntamos si se generaría una señal de potencial evocado (ERP) asociado a la aparición de una incongruencia entre el movimiento de la escena y el término verbal mostrado (incongruencia de tipo semántica). En experimentos con lenguaje natural, la aparición de una palabra semánticamente incongruente dentro de una frase genera una negatividad con topografía centroparietal que alcanza su máximo alrededor de los 400 ms luego de la aparición de la palabra incongruente (Kutas y Hillyard, 1980; Hahne y Friederici, 2002), denominándose por ello N400.

Materiales y métodos

Experimento 1












Participantes

Diecinueve voluntarios sanos ($27,2 \pm 0,7$ años). Todos eran diestros, hablaban español como primera lengua y tenían estudios terciarios en curso o completos. Los sujetos no tenían antecedentes de desórdenes neurológicos o psiquiátricos, ni se encontraban bajo medicación alguna al momento del experimento.

Materiales

Se realizó una modificación del protocolo de Fernández Feijóo *et al.* (2008, aumentándose el número de estímulos. Se crearon 150 escenas visuales diferentes (duración: $3,34 \pm 0,01$ s), con las combinaciones de 5 figuras geométricas y 6 movimientos distintos (Tabla 1), y se crearon 150 frases que describieran las distintas escenas. Cada frase seguía el patrón *BAPA-Sujeto-Verbo-NALA-Objeto*, en donde tanto el sujeto como el objeto de cada frase referían a una de las figuras geométricas, y el verbo refería al movimiento en la escena.

Tabla 1. Estímulos utilizados para la confección de las frases de entrenamiento y testeo.

ELEMENTO/ACCIÓN	PALABRA	SIGNIFICADO
	Bapa	"El"
	Nala	"Hacia el"
	Bare	
	Lane	
	Fadi	
	Mere	
	Pefa	
	Siru	"Acercarse"
	Revu	"Alejarse"
	Fenu	"Pasar horizontalmente"
	Melu	"Pasar verticalmente"
	Naru	"Pasar en diagonal"
	Jazu	"Girar alrededor"

Procedimiento

Entrenamiento

Cada participante se encontraba sentado en un cuarto aislado eléctrica y acústicamente, frente a un monitor de computadora de 17 pulgadas, a una distancia de aproximadamente 0,5 m. El entrenamiento comenzaba con una pantalla donde se indicaba que el objetivo del experimento era aprender el significado de las distintas palabras de un idioma inventado. Para ello, se presentarían distintas escenas junto con frases que las describían en dicho idioma. Luego, cada ensayo consistía en la presentación de una escena diferente, y simultáneamente una frase que describía la escena. La frase aparecía en formato audiovisual por debajo de cada escena, palabra por palabra de izquierda a derecha, con un intervalo entre palabras de 350 ms. En esta etapa se presentaron 70 ensayos diferentes, cada uno de ellos dos veces consecutivas. El intervalo entre ensayos fue de 3.000 ms.

Para asegurar que los sujetos prestaran atención, cada 4 a 7 frases se les preguntaba en formato visual si cierta palabra había aparecido en la frase previa. Para responder, debían apretar Ctrl derecho para “Sí”, y Ctrl izquierdo para “No”.

Testeo

Completado el entrenamiento, una pantalla indicaba que se testearía si el participante había aprendido el significado de las diferentes palabras. Luego, cada ensayo consistía en una escena nueva junto con una frase nueva en formato audiovisual. Solo en el 50% de los casos la frase describía la escena mostrada (frase Correcta), mientras que en el restante 50% el “verbo” no correspondía al movimiento mostrado en la escena visual (frase Incorrecta). Los sujetos debían decidir lo más rápidamente posible si la frase describía la escena o no, presionando en caso afirmativo Ctrl derecho, y en caso negativo Ctrl izquierdo. El tiempo límite para responder eran 2.000 ms luego de la aparición completa de

cada frase. En total se presentaron las 80 escenas no vistas durante el entrenamiento, cada una de ellas una sola vez. En simultáneo se registraron los electroencefalogramas de los participantes. Se registraron además el porcentaje de respuestas, la tasa de aciertos y el tiempo de respuestas de cada individuo.

Registro EEG

El electroencefalograma fue registrado a 256 HZ mediante un equipo Akonic Bio-PC de 19 electrodos (Sistema Internacional 10/20, referencia biauricular). La resistencia de los electrodos fue mantenida por debajo de 10 k Ω durante todos los registros. Cada registro fue filtrado *off line* de 1 a 20 Hz y rerreferenciado a la referencia promedio de los 19 electrodos. Luego fue separado en épocas por condición experimental (Correctas-Incorrectas), abarcando 200 ms (línea de base) y 1.800 ms posteriores a la aparición del “verbo” de cada frase. Se analizaron solamente los ensayos en los cuales los sujetos respondieron acertadamente (aciertos). Las épocas fueron examinadas visualmente y excluidas del análisis si contenían artificios oculares o de otro tipo. De esta manera, se excluyó alrededor del 5% de las épocas.

Estadística

Para establecer quiénes respondieron acertadamente por encima del azar (50%) se realizó un test de G de bondad de ajuste (Zar, 1999) con el número de aciertos y desaciertos por individuo. Las respuestas de los individuos que no superaron el criterio fueron descartadas para los análisis posteriores. Para el resto, se discriminó por tipo de frase (Correcta o Incorrecta) el porcentaje de respuesta, el de aciertos y el tiempo de respuesta promedio de los aciertos. Los datos individuales fueron ingresados en un análisis poblacional, realizándose un test de *t* pareado para cada una de estas tres variables.

Resultados

En total 15 de los 19 participantes respondieron significativamente por encima del azar durante la fase de testeo (test de G de bondad de ajuste, $p < 0,05$). El porcentaje de aciertos para este grupo fue de $87,00 \pm 2,54\%$, siendo el promedio de respuestas efectuadas de un $78,00 \pm 5,33\%$. El porcentaje de respuestas y el de aciertos fue mayor para las frases Correctas que para las Incorrectas ($t_{14} = -2,45$; $p < 0,05$ y $t_{14} = -3,61$; $p < 0,005$, respectivamente). El tiempo de respuesta no difirió entre ambos tipos de frases ($t_{14} = -1,29$; $p = 0,22$), aunque se observó una tendencia a mayores tiempos de respuesta en las frases Correctas comparadas con las Incorrectas.

La inspección visual del Gran Promedio mostró diferencias entre condiciones en los canales frontales y parieto-occipitales entre los 400 y 700 ms luego de la presentación del “verbo” (Figura 1). Por tanto, se definieron dos ventanas para el análisis estadístico: 0-400 ms y 400-700 ms, luego de la aparición del “verbo”. Para cada ventana temporal se calculó el voltaje promedio por condición y participante para cada uno de los 19 electrodos de registro utilizados, los cuales fueron ingresados en un Anova de medidas repetidas con los factores Electrodo (Fp1 a Pz) y Condición (Correctas e Incorrectas). El ajuste de Greenhouse-Geisser fue aplicado para corregir las violaciones a la esfericidad.

El análisis de los datos de los primeros 400 ms posteriores a la aparición del verbo no mostró efectos principales de Condición ($F_{1,14} = 0,14$, $p = 0,71$), ni una interacción significativa entre Electrodo y Condición ($F_{18,252} = 0,694$, $p = 0,562$). La ventana de 400 a 700 ms mostró una interacción significativa entre Electrodo y Condición ($F_{18,252} = 5,016$; $p < 0,005$). Mientras en la condición Correctas la presentación de frases que concordaban con la escena presentada elicita una negatividad fronto-central, la incongruencia entre el movimiento de la escena y el verbo elicita una negatividad centro-parietal

característica de un componente de tipo N400 (Figura 1). Las comparaciones *post hoc* revelaron diferencias distribuidas en los electrodos frontales y parieto-occipitales ($p < 0,05$).

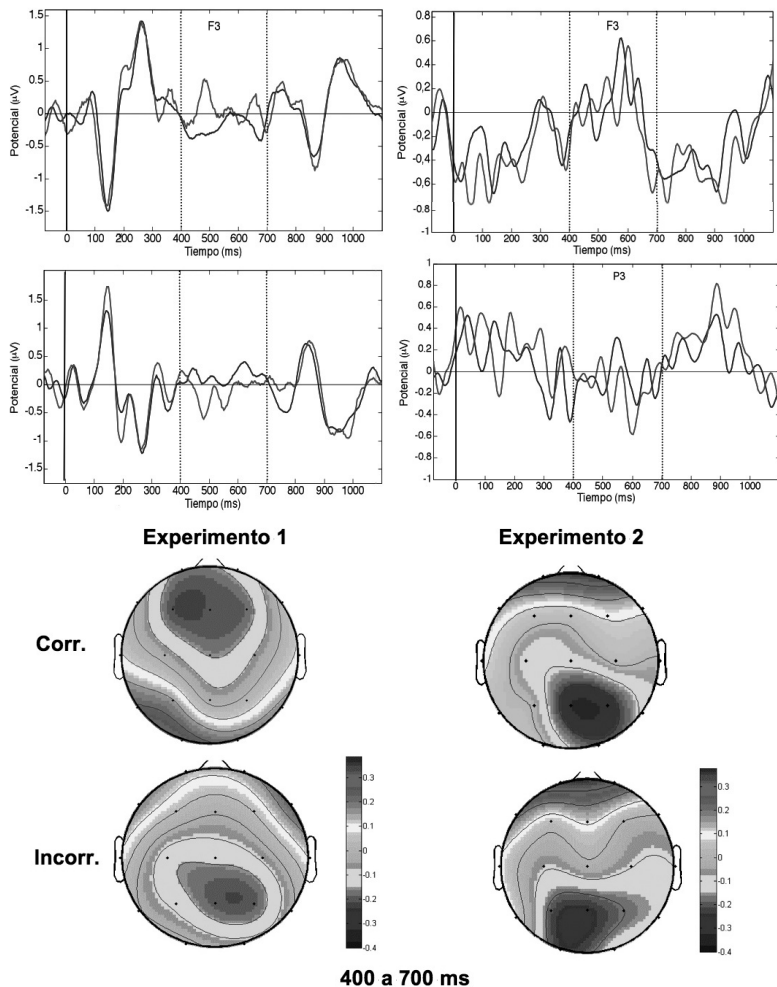


Figura 1. Panel Superior y Medio. Gran Promedio para los Canales F3 y P3 en los Experimentos 1 y 2 para las condiciones Correctas e Incorrecta. Las líneas punteadas verticales indican la ventana de 400 a 700 ms desde la aparición del verbo. Panel Inferior. Topografías en ambos experimentos para la condición Correcta (Corr.) e Incorrecta (Inc.) en la ventana de 400 a 700 ms luego de presentado el verbo.

Experimento 2

Materiales y métodos

Participantes

Veinticuatro voluntarios adultos sanos de entre 20 y 35 años (media $25,7 \pm 0,9$ años). Todos los sujetos cumplieron las mismas condiciones que para el experimento anterior.

Estímulos

Se utilizó la misma lista de ensayos que para el experimento 1.

Procedimiento

Entrenamiento

El entrenamiento tuvo la misma serie de ensayos que el experimento 1, pero las frases aparecieron solo en formato auditivo. El resto de los parámetros de presentación de cada ensayo fue idéntico al del primer experimento.

Testeo

El testeo se realizó con la misma serie de ensayos que en el experimento 1, pero las frases se presentaron solo en formato auditivo. El resto de los parámetros de presentación de cada ensayo fue idéntico al del primer experimento. Se tomaron las mismas medidas comportamentales que para el primer experimento, junto con el registro electroencefalográfico.

EEG

El protocolo de análisis de los EEG fue idéntico al del experimento 1.

Estadística

Los análisis fueron realizados de igual manera que para el primer experimento.

Resultados

De los 24 participantes, 16 respondieron por encima del azar durante la fase de testeo (ca. 67%) (test de G de bondad de ajuste, $p < 0,05$). Para este grupo, el porcentaje de respuestas fue $76,67\% \pm 4,92\%$, y el de aciertos, $86,41 \pm 2,49\%$. El porcentaje de respuestas y de aciertos fue mayor para las frases Correctas que para las Incorrectas ($t_{14} = -2,40$; $p < 0,05$ y $t_{14} = -2,67$; $p < 0,05$, respectivamente). El tiempo de respuesta fue mayor en las frases Correctas comparadas con las Incorrectas ($t_{14} = -3,766$; $p < 0,005$).

La inspección visual del Gran Promedio no mostró diferencias entre la presentación de frases Correctas e Incorrectas en ningún momento de los trazados. Al igual que para el experimento 1, en la ventana de 400 a 700 ms posteriores a la aparición del “verbo” se halló una negatividad parieto-occipital, pero esta vez tanto para la presentación de frases Correctas como Incorrectas (Figura 1). El Anova de medidas repetidas no mostró un efecto principal de Condición ($F_{1,14} = 0,175$; $p = 0,682$) ni una interacción significativa entre Electrodo y Condición ($F_{18,252} = 0,519$; $p = 0,733$).

Discusión

Los resultados obtenidos en el presente trabajo extienden los registrados por Fernández Feijóo *et al.* (2008). En ambos experimentos la mayoría de los participantes pudo inferir el significado de los 6 términos verbales nuevos mediante el seguimiento de la coocurrencia entre su aparición y la información contextual de las escenas mostradas en el entrenamiento, aun en ausencia de conocimiento previo del “idioma” mostrado. Al igual que en Fernández Feijóo *et al.* (2008), los porcentajes de respuestas y de aciertos fueron

significativamente mayores para las frases Correctas comparados con las de Incorrectas. Esto indica que los participantes superaron el límite para responder más veces bajo la condición Incorrectas. Siendo que en cada ensayo la aparición de la escena precedía la del lexema “verbal” (1335 ± 27 ms) y puesto que el grado de aprendizaje no fue perfecto, es posible que la visualización del movimiento activara varios candidatos lexicales previo a la aparición del lexema escrito. La correspondencia entre este lexema y una de las formas activadas llevaría al individuo a decidir que la frase es correcta, mientras que la falta de correspondencia entre ambos llevaría a la decisión de que la frase es incorrecta. Esta falta de coincidencia entre lo presentado y lo esperado podría disparar procesos de recomparación entre la/s forma/s lexical/es activada/s y el lexema escrito, aumentando el tiempo de decisión durante los ensayos de la condición Incorrectas.

A su vez, el mayor porcentaje de aciertos en la condición Correctas indica que la probabilidad de decidir que una frase Incorrecta era correcta fue mayor que la probabilidad de decidir que una frase Correcta era incorrecta. Si la aparición del movimiento en la escena activara más de un candidato lexical, es probable que uno de ellos fuera el correcto y el resto no. Al presentar una frase correcta, habría entonces una alta probabilidad de coincidencia entre uno de los candidatos activados y el presentado, siendo así alta la probabilidad de que el individuo acertara la respuesta. Al presentar una frase incorrecta, si el lexema de la frase coincidiera con uno de los candidatos lexicales incorrectos activados, el sujeto podría decidir que la frase es correcta aun cuando no lo fuera, existiendo una mayor probabilidad de dar como correcta una frase incorrecta.

Por otra parte, tanto en el experimento 1 como en el 2 se observaron mayores tiempos de respuesta en la condición Correctas. Aunque esta diferencia fue significativa solo en el experimento 2, la tendencia fue clara también en el primer

experimento. Dado que a los participantes no se les instruyó que los errores de las frases estarían en el verbo, el resultado podría estar relacionado con que, durante la presentación de una frase Correcta, el sujeto debería esperar necesariamente hasta la compleción de la frase para poder decidir que no se presentaban errores, mientras que durante la presentación de una frase incorrecta la aparición del error en la posición del verbo podría disparar el proceso de decisión antes que en el caso de las frases correctas.

Al comparar ambos experimentos mediante un Anova de medias repetidas (Condición intrasujeto, experimento entre sujetos) no se halló efecto Principal o interacción Experimento x Condición ni para el Porcentaje de respuestas ni para el de aciertos, aunque se observó una tendencia a mayor porcentaje de respuestas y aciertos para el experimento 1 respecto del experimento 2.

En el caso del tiempo de respuesta hubo un efecto principal de experimento ($F_{1,28} = 4,422$; $p < 0,05$), siendo mayores para el segundo experimento en comparación con el primero. Estos resultados podrían indicar que la presentación escrita y auditiva de la frase facilitó la realización de la tarea en el experimento 1.

Con respecto al registro de EEG, en el experimento 1 se hallaron diferencias en los potenciales generados entre los 400 y los 700 ms luego de la aparición del lexema “verbal”. Mientras que en la condición Correctas se obtuvo una negatividad fronto-central, en los ensayos de Incorrectas se generó una negatividad parieto-occipital de tipo N400. La negatividad frontal observada en los ensayos de Correctas podría estar relacionada con procesos atencionales dirigidos específicamente hacia el movimiento desplegado en la escena, disparados por la aparición del término verbal en la frase. La latencia y ubicación frontal del componente negativo encontradas son similares al componente ADAN (Attention Driven Anterior Negativity), el cual se elicitaba típicamente ante la aparición de

un estímulo clave (“cue”) que señala la aparición posterior de un estímulo relevante para la tarea (Eimer *et al.*, 2002; Seiss *et al.*, 2007). Si bien este componente ha sido descripto típicamente como de distribución contralateral a la posición en la que se espera la aparición del estímulo relevante, es posible que en el presente trabajo, dada la ubicación siempre central en la pantalla de la computadora del movimiento desplegado en cada ensayo, se generara una distribución más bilateral del componente.

A su vez, la generación de un componente de tipo N400 ante la aparición del término verbal incorrecto muestra que la incongruencia entre el movimiento de la escena y el lexema presentado genera un componente cognitivo similar al observado en protocolos psicolingüísticos clásicos. Su aparición en los ensayos de Incorrectas apoya una interpretación incremental de la frase con una integración semántica temprana de la información contextual (Sedivy *et al.*, 1999; Knoeferle *et al.*, 2009). Knoeferle *et al.*, (2009) realizaron un experimento con registro EEG, en el cual presentaron 160 dibujos diferentes, cada uno de los cuales representaba una acción distinta. Luego de la exposición a cada imagen (por al menos 3 s), esta desaparecía, y se presentaba, palabra por palabra, una frase escrita describiendo la imagen previa o con un error en el verbo respecto de la acción de la imagen. Los resultados mostraron la generación de un componente de tipo N400 ante la aparición de un verbo incongruente con la acción presentada en la imagen. Los autores interpretaron sus resultados como indicativos de un proceso incremental de interpretación de la información semántica de la frase, que integra y verifica de manera continua la información contextual y lingüística. Siendo que en el presente trabajo escena y frase fueron presentados simultáneamente, la generación de un N400 ante la aparición de los lexemas verbales incongruentes con el movimiento de la escena indicaría que el proceso de integración y verifica-

ción de la información lingüística con la contextual es aún más rápido que lo pensable por medio de paradigmas de presentación de imagen y frase no simultáneos.

Los registros del experimento 2 mostraron también la aparición de una negatividad centro-parietal entre los 400 y 700 ms luego de la aparición del lexema verbal, pero en ambas condiciones experimentales. La aparición de una negatividad centro-parietal en la condición Correctas podría estar relacionada con un mayor grado de incertidumbre del participante en este experimento, lo cual podría deberse a una mayor dificultad en la tarea, relacionada probablemente con un menor grado de aprendizaje. Los resultados comportamentales apuntan a una mayor dificultad para la tarea en el experimento 2 comparado con el experimento 1.

En conclusión, los resultados obtenidos hasta el momento muestran que en adultos es posible inferir el significado de términos verbales nuevos mediante el seguimiento explícito de la coocurrencia entre término y contexto visual, y que la integración de ambas informaciones parece darse de manera simultánea. Como próximo paso nos planteamos estudiar si es posible la adquisición de términos verbales nuevos mediante el aprendizaje de tipo implícito.

Bibliografía

- D'Odorico, L. y Fasolo, M. 2007. "Nouns and verbs in the vocabulary acquisition of Italian children", *Journal of Child Language*, 34, pp. 891-907.
- Eimer, M.; Van Velzen, J. y Driver, J. 2002. "Crossmodal interactions between audition, touch and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations", *Journal of Cognitive Neuroscience*, 14, pp. 254-271.
- Fernández Feijóo, M. E.; Vernis, S.; Zanutto, S. y Wainseboim, A. 2008. "Adquisición de significados verbales nuevos por un proceso de asociación multimodal", *Revista Argentina de Neuropsicología*, 12, pp. 1-12.

- Gentner, D. 1982. "Why nouns are learned before verbs: Linguistic relativity versus natural partitioning", en Kuczaj, S. (ed.). *Language development* (2), *Language, thought and culture*. New Jersey, Lawrence Erlbaum Associates, pp. 301-334.
- Hahne, A. y Friederici, A. D. 2002. "Differential task effects on semantic and syntactic processes as revealed by ERPs", *Cognitive Brain Research*, 13, pp. 339-356.
- Knoeferle, P.; Urbach, T. P. y Kutas, M. 2009. "Is incremental semantic interpretation related to end-of-sentence verification?: Evidence from correlation analyses", en Taatgen, N. A. y van Rijn, H. (eds.). *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, pp. 1127-1132.
- Kutas, M. y Hillyard, S. A. 1980. "Reading senseless sentences: brain potentials reflect semantic incongruity", *Science*, 207, pp. 203-205.
- Lacerda, F. y Sundberg, U. 2006. "An Ecological Theory of Language Acquisition", *Revista de Estudos Linguísticos da Universidade do Porto* 1, pp. 53-106.
- Marcus, G.; Vijayan, S.; Byi Rao, S. y Vishton, P. M. 1999. "Rule learning by seven-month-old infants", *Science* 283, pp. 77-80.
- Marklund, E. y Lacerda, F. 2006. "Infants' ability to extract verbs from continuous speech", *Ninth International Conference on Spoken Language Processing*, pp. 1403-1406.
- Saffran, J. R.; Aslin, R. N. y Newport, E. L. 1996. "Statistical learning by 8-month-olds infants", *Science* 274, pp. 1926-1928.
- Sedivy, J. C.; Tanenhaus, M. K.; Chambers, C. G. y Carlson, G. N. 1999. "Achieving incremental semantic interpretation through contextual representation", *Cognition* 71, pp. 109-148.
- Seiss, E.; Gherri, E.; Eardley, A. F. y Eimer, M. 2007. "Do ERP components triggered during attentional orienting represent supramodal attentional control?", *Psychophysiology*, 44(6), pp. 987-990.
- Siskind, J. 1996. "A computational study of cross-situational techniques for learning word-to-meaning mappings", *Cognition*, 61, pp. 39-61.
- Smith, L. y Yu, C. 2008. "Infants rapidly learn word-referent mappings via crosssituational statistics", *Cognition*, 106(3), pp. 1558-1568.
- Yu, C. y Ballard, D. 2007. "A unified model of early word learning: Integrating statistical and social cues", *Neurocomputing*, 70, pp. 2149-2165.
- Yu, C. y Smith, L. B. 2007. "Rapid word learning under uncertainty via cross-situational statistics", *Psychological Science*, 18(5), pp. 414-420.
- Zar, J. H. 1999. *Biostatistical analysis*, New Jersey, Prentice Hall.