

Poca evidencia de un sesgo sustantivo en el habla espontánea tseltal**

Marisa Casillas
The University of Chicago, USA

Ruthe Foushee
The New School for Social Research, USA

Juan Méndez Girón
Centro de Investigación y Estudios Superiores en Antropología Social Sureste, Mexico

Gilles Polian
Centro de Investigación y Estudios Superiores en Antropología Social Sureste, Mexico

Penelope Brown
Max Planck Institute for Psycholinguistics, The Netherlands

Autora correspondiente:

Marisa Casillas
Department of Comparative Human Development,
Rosenwald Hall 3rd Floor,
1101 East 58th Street,
Chicago, IL 60637, USA.
Email: mcasillas@uchicago.edu

Abstracto

Este estudio examina si los niños que adquieren tseltal (maya) demuestran un sesgo sustantivo: una representación excesiva de los sustantivos en los vocabularios tempranos de los niños. Se argumenta que los sustantivos, específicamente los sustantivos concretos y animados, predominan universalmente en los vocabularios tempranos de los niños porque sus referentes están naturalmente disponibles como conceptos acotados a los que se pueden asignar etiquetas lingüísticas. Esta ventaja temprana para el aprendizaje de sustantivos se ha documentado utilizando múltiples métodos y en una colección diversa de poblaciones lingüísticas. Sin embargo, la evidencia anterior sobre un sesgo sustantivo entre los estudiantes de tseltal ha sido contradictoria. Las características gramaticales tseltales y los patrones de interacción niño-cuidador amortiguan la prominencia de los sustantivos y aumentan la prominencia de los verbos, lo que lleva a la predicción de una disminución del sesgo de los sustantivos y tal vez incluso un predominio temprano de los verbos. Aquí analizamos el uso de raíces nominales y verbales en el habla espontánea de los niños a partir de grabaciones egocéntricas de un día de duración.29 aprendices tseltales entre 0;9 y 4;4. Encontramos

** Este manuscrito fue traducido del inglés con la ayuda de [Google Translate](https://www.google.com/translate).

evidencia débil o nula de un sesgo sustantivo utilizando dos enfoques analíticos separados sobre los mismos datos; un análisis arroja una sugerencia preliminar de un resultado invertido (es decir, un sesgo verbal). Discutimos las implicaciones de estos hallazgos para teorías más amplias sobre el sesgo de aprendizaje en el desarrollo léxico temprano.

Palabras clave: Tseltal, desarrollo léxico, vocabulario, sesgo sustantivo, grabaciones diurnas

1. Introducción

Las palabras para objetos predominan en los vocabularios tempranos de los niños, especialmente palabras para objetos concretos y nombres de personas (Genter, 1982; 2006; Gentner y Boroditsky, 2001; Gleitman, 1990; Nelson, 1973). Esta tendencia, denominada “sesgo sustantivo”, se ha propuesto como una característica universal del desarrollo temprano del lenguaje, que da a los niños un pie en la puerta hacia partes más complejas del sistema lingüístico (Au, Dapretto y Song, 1994; Bassano, 2000; Caselli et al., 1995; Gentner & Boroditsky, 2001; 2008; Jackson-Maldonado et al., 1993; Tardif et al., 1997; 1999; ver también Baillargeon, 1993; Soja et al., 1991; Spelke, 1985; 1993). La Hipótesis de Particiones Naturales/Relatividad Relacional de Gentner (NP/RR; 1982; 2006) propone que el sesgo sustantivo surge porque los objetos se dan perceptivamente como conceptos individuables, incluso para los bebés. Por el contrario, los conceptos relacionales (como las acciones sobre objetos) deben seleccionarse lingüísticamente antes de que puedan alcanzar la individualidad. Desde este punto de vista, los objetos concretos son conceptualmente frutos maduros, listos para ser asignados a las etiquetas lingüísticas correspondientes. Mientras tanto, los conceptos relacionales tienden a estar disponibles más gradualmente y, por lo tanto, sólo más adelante estarán listos para una etiqueta lingüística correspondiente.

Las investigaciones anteriores que examinan el sesgo del sustantivo generalmente han adoptado uno de tres enfoques:

1. **Aprendizaje de palabras novedosas en un contexto experimental:** Podemos examinar las condiciones bajo las cuales los niños aprenden nuevos objetos y palabras de acción. Por ejemplo, los niños pequeños estadounidenses esperan que las palabras novedosas se refieran a objetos más que a acciones (Waxman, 1991; 1993). También aprenden nuevos sustantivos más fácilmente que nuevos verbos, incluso si los niños están muy atentos a las nuevas acciones demostradas (Childers et al., 2020; 2022; Imai et al., 2008);
2. **Proporción o ratio de tipos sintácticos en el vocabulario infantil:** Un enfoque más tradicional es examinar el desarrollo natural del vocabulario de los niños. Por ejemplo, en seis idiomas, Gentner (1982) encontró que los sustantivos constituían entre el 50% y el 85% de los tipos de palabras en los vocabularios productivos tempranos de los niños, mientras que los predicados (incluidos los verbos) sólo representaban entre el 0% y el 35%. Desde entonces, muchos otros estudios han seguido su ejemplo (p. ej., Setoh et al., 2021; Tardif et al., 1997; 1999). En algunos casos, los investigadores basaron estos cálculos en el habla infantil espontánea, pero en otros casos los cálculos se basaron en el vocabulario infantil informado por los padres;
3. **Representación relativa de tipos sintácticos en el vocabulario infantil:** Inspirándose en los análisis de Bates y colegas (1994) del inventario de Desarrollo Comunicativo de MacArthur Bates (“CDI”, una lista de verificación de vocabulario; Bates et al., 1994; 1995; Fenson, 2007), un tercer enfoque ha sido probar si una palabra de la clase está sobrerrepresentada o subrepresentada en relación con el tamaño total del vocabulario de un niño. Capta la intuición de que si un niño ha adquirido el 10% de su vocabulario

general pero ha adquirido el 20% de sus sustantivos, los sustantivos están relativamente sobrerrepresentados. Utilizando este enfoque, un trabajo reciente de Braginsky, Frank y sus colegas encontró un sesgo sustantivo constante en los vocabularios productivos de los niños informados por los padres en 27 idiomas (Braginsky et al., 2019; Frank et al., 2021).

En el presente artículo nos centramos en enfoques no experimentales, entre los cuales el segundo y tercer enfoque tienen fortalezas complementarias. El segundo enfoque (medidas de frecuencia proporcional) puede basarse en datos del habla espontánea altamente naturalistas, listas de verificación de vocabulario o una combinación, pero no mide directamente la representación excesiva o insuficiente de las clases de palabras. El tercer enfoque (representación relativa) ofrece un método claro para medir la sobre y la subrepresentación, pero sólo se ha intentado con datos de vocabulario informados por los padres, lo que limita su validez ecológica y aplicabilidad interlingüística (las listas de verificación existen en un formato más amplio). número limitado de lenguas que los corpus de habla espontánea).

En el presente estudio utilizamos ambos enfoques con datos del habla espontánea para examinar la evidencia de un sesgo sustantivo en niños que adquieren tseltal (geolecto de Tenejapan; maya; Chiapas, México). Como veremos más adelante, la universalidad propuesta de un sesgo de objeto conceptual, si bien está bien motivada, se ve socavada por datos limitados sobre lenguas no indoeuropeas, especialmente lenguas de menor escala. Argumentamos que el tseltal es un caso de prueba particularmente informativo para la universalidad del sesgo sustantivo, considerando que tanto factores lingüísticos como culturales pueden llevar a los niños tseltales a adquirir verbos tan fácilmente como los sustantivos, sino más fácilmente. A continuación, revisamos brevemente la evidencia previa sobre el sesgo universal de los sustantivos, con especial atención a trabajos previos sobre lenguas mayas (incluido el tseltal), antes de profundizar en los métodos y hallazgos del presente estudio.

1.1 El sustantivo sesgo: una breve descripción

La evidencia translingüística generalmente favorece el sesgo sustantivo, pero cuenta una historia complicada. Se ha encontrado evidencia a favor de un sesgo sustantivo en los vocabularios tempranos productivos de los niños en una variedad de idiomas: la mayoría son indoeuropeos¹—incluidos inglés*, alemán*, sueco*, noruego*, danés*, italiano*, francés*, español*, portugués*, griego*, croata*, letón*, checo*, ruso*, eslovaco*, turco, hebreo. , japonés, coreano, mandarín, cantonés, kiswahili, kigiriamay Kaluli (ver, por ejemplo, Braginsky et al., 2019; Frank et al., 2021; Gentner, 1982; Imai et al., 2008; Snedeker et al., 2003; Setoh et al., 2021; Pae, 1993; Tardif et al. otros, 1997; 1999). Entre los estudios que comparan directamente el aprendizaje de palabras en dos o más idiomas, se ha encontrado evidencia parcial de un sesgo sustantivo utilizando los tres tipos de datos mencionados anteriormente: aprendizaje de palabras basado en experimentos (Snedeker et al., 2003; Imai et al., 2008) , producción de habla naturalista (Gentner, 1982; Setoh et al., 2021; Tardif et al., 1997; 1999) y datos de listas de verificación de vocabulario (Gentner, 1982; Tardif et al., 1999; Braginsky et al., 2019). Es importante destacar que estos estudios comparativos plantean la posibilidad de

¹ Las lenguas indoeuropeas (marcadas con un asterisco arriba) son habladas por casi la mitad de la población mundial, pero sólo representan alrededor del 6% de las ~7000 lenguas del mundo (Skirgård, 2017; véase también Hammerström et al. (2023) y Simons y Fennig (2017) para más información). Seis de los nueve idiomas no indoeuropeos de esta lista son idiomas igualmente importantes que se hablan en grandes sociedades postindustriales. La lista se diversifica un poco si incluimos estudios de palabras. *comprensión* (p. ej., ver Braginsky et al., 2019; Frank et al., 2021).

que, incluso si el sesgo sustantivo es universal, surge más débilmente en algunas lenguas y con más fuerza en otras.

Entre las lenguas no indoeuropeas, el mandarín y el coreano han sido estudiados con mayor frecuencia por un sesgo sustantivo, con trabajos que utilizan listas de verificación de vocabulario, estudios de diarios, habla espontánea y experimentos, y con diseños de estudio que tienen muestras de edad transversales o longitudinales. de niños pequeños (p. ej., ambos idiomas: Braginsky et al., 2019; Frank et al., 2021; y mandarín: Setoh et al., 2021; Tardif, 1996; Tardif et al., 1997, 1999, 2008; Xuan & Dollaghan, 2013; y coreano: Au et al., 1994; Bornstein et al., 2004; Chang-Song, 1997; Choi, 2000; Choi y Gopnik, 1995; Gopnik y Choi, 1990; Gopnik et al., 1996; Kim et al., 2000; Pae, 1993). Entre estos estudios hay evidencia a favor y en contra del sesgo sustantivo. En algunos casos, los hallazgos respaldan un predominio temprano de sustantivos, pero también respaldan una ventaja relativa para aprender verbos, en comparación con los angloparlantes o con el vocabulario en inglés de los mismos niños (este último para bilingües inglés-mandarín; Setoh et al., 2021 ; Xuan y Dollaghan, 2013).

Entre las poblaciones de lenguas indígenas y de menor escala, sólo conocemos evidencia que se relaciona directamente con el debate sobre el sesgo sustantivo de tseltal (maya), tsotsil (maya), Navajo (Athabaskan-Eyak-Tlingit) y Kaluli (Trans-Nueva Guinea). Con dos excepciones: Brown, Braun y Gentner (2005) y Gentner y Boroditsky (2008)—Los datos de estas lenguas provienen de transcripciones de habla espontánea. Los datos navajos se recopilaron como vocabulario informado por los cuidadores a través de una adaptación del CDI realizada por un equipo de hablantes y académicos navajos (Gentner y Boroditsky, 2008) y se encontró que apoyaban un sesgo sustantivo. Los datos de Kaluli, recopilados originalmente por Schieffelin (1979) fueron analizados posteriormente por Gentner (1982) a través de traducciones al inglés de las transcripciones originales; favoreció un sesgo sustantivo (Gentner, 1982; 2006). Los datos de las dos lenguas mayas (discutidos más adelante) argumentan en contra de un sesgo sustantivo, a favor de un sesgo verbal (Brown, 1998, 2008; de León, 1999a, 1999b, 2001; pero véase la discusión a continuación). Entonces, en general, entre las lenguas indígenas y de menor escala, la evidencia de un sesgo sustantivo es mixta.

1.2 El sustantivo parcialidad en las lenguas mayas

El desarrollo del vocabulario maya ha jugado un papel especial en este debate sobre el sesgo sustantivo; ha aportado algunos de los únicos análisis lingüísticos detallados de comunidades lingüísticas de pequeña escala, motivados por factores tanto culturales como lingüísticos que pueden promover el aprendizaje temprano de los verbos. La evidencia de dos lenguas mayas en particular, el tseltal y el tsotsil, ha hecho contribuciones esenciales a este debate (de León, 1999a; 1999b, 2001; Brown, 1997; 1998; 2001; 2007; 2008; Brown et al., 2005; Pye et otros, 2007). Las normas de interacción de la comunicación niño-cuidador y la estructura lingüística de ambos idiomas aumentan la prominencia de los verbos y disminuyen la prominencia de los sustantivos en el aporte lingüístico de los niños. Por ejemplo, en tseltal, el orden básico de las palabras verbo-objeto-sujeto (VOS) coloca a los verbos en una posición muy destacada. Además, los argumentos que tienen referencias cruzadas sobre el verbo se descartan habitualmente, de modo que muchas expresiones comprenden un solo verbo flexionado. En términos de complejidad morfosintáctica, los sustantivos y los verbos están en igualdad de condiciones que en inglés y otros idiomas, porque la morfología flexiva obligatoria de los sustantivos y los verbos es muy regular y fonológicamente similar, lo que permite a los niños desarrollar una estrategia de segmentación morfológica que se aplica bien a ambos tipos de palabras. También hay muchos verbos “específicos” de uso frecuente: estos son verbos transitivos que están muy restringidos en cuanto a los pacientes que toman y, por lo tanto, son muy concretos en su semántica (p. ej., tseltalde’ para ‘comer/morder-[carne]’). Por el contrario,

** Este manuscrito fue traducido del inglés con la ayuda de [Google Translate](https://www.google.com/translate/).

los sustantivos inanimados tienden a referirse a material no individualizado, lo que va en contra del mapeo simple e individualizado de objeto a etiqueta asumido como parte de la hipótesis NP/RR. Finalmente, se observa que los cuidadores adultos rara vez etiquetan objetos para niños, más a menudo hablan de contexto relacional o comentan sociointeractivamente sobre el comportamiento del niño (Brown, 1998; 2011; 2014; de León, 1999a, 199b, 2001).

De hecho, la investigación longitudinal y en profundidad de Brown (1998) sobre el habla espontánea temprana de dos niños encontró evidencia de una facilitación del aprendizaje temprano de los verbos y ninguna evidencia de un sesgo temprano de los sustantivos. Examinó el habla espontánea de ambos niños a partir de grabaciones de audio de 2 horas mensuales y grabaciones de video adicionales cada seis semanas durante el período de muestreo (edades 1;3–2;3 y 1;5–2;5). En las primeras etapas del desarrollo morfosintáctico (MLU 1.0-1.5), el habla de los niños reveló inventarios de sustantivos y verbos de tamaño similar. Con el tiempo, los inventarios de verbos de los niños comenzaron a superar a sus inventarios de sustantivos (terminan en MLU = 1,7; 54–59% de verbos; 53–57% de verbos cuando se incluyen nombres propios).² También descubrió que, entre las expresiones multimorfémicas de los niños, las expresiones verbales (por ejemplo, verbos flexionados) superan a las expresiones nominales (sustantivos flexionados, sustantivos compuestos). Luego, profundizando más en su adquisición morfosintáctica, descubrió que, al final del período de observación (MLU = 1,7), los niños tenían un dominio productivo parcial de algunos prefijos ergativos y sufijos absolutivos, que se aplican tanto a sustantivos como a verbos. Esto sugiere que los niños tseltal en las primeras etapas de la producción del habla ya han comenzado a desarrollar sistemas morfosintácticos abstractos que pueden ayudar simultáneamente en el aprendizaje de sustantivos y verbos. Sólo en los verbos, los niños también mostraron evidencia temprana del sufijo benéfico y de la marca de aspecto obligatorio y no obligatorio, pero aún no de manera sistemática. En particular, muchos de los verbos de los niños eran verbos específicos, es decir, verbos transitivos que están muy restringidos en los pacientes que atienden. Entonces, en pocas palabras, en el habla espontánea temprana de dos niños, no encontró ningún sesgo nominal inicial (excluyendo los nombres propios), luego un rápido crecimiento en el inventario de verbos y evidencia de uso de morfología flexiva que sugiere un aprendizaje morfosintáctico abstracto temprano. Estos datos sugieren una capacidad temprana sorprendente para el aprendizaje de verbos en adquiriendo tseltal niños.

Estos hallazgos se ven reforzados por patrones cualitativamente similares en los análisis longitudinales en profundidad de la educación temprana de dos niños. Sobre Tsotsi producciones (MLU 1–2.75; de León, 1999a, 1999b, 2001). Por ejemplo, los sustantivos nunca superaron la marca del 50% en la Sobre Tsotsi Los primeros vocabularios productivos de los niños (incluso cuando incluían nombres propios y términos de parentesco) y las expresiones multimorfémicas tenían muchas más probabilidades de ser expresiones verbales que nominales. Dado que Sobre Tsotsi es una lengua maya estrechamente relacionada con el tseltal y comparte muchos de los mismos factores lingüísticos e interaccionales que apoyan el aprendizaje temprano de los verbos, podemos tratarlos Sobre Tsotsi hallazgos como una segunda fuente muy cercana de datos sobre la existencia de un sesgo sustantivo tseltal. Juntos, los dos estudios forman un caso sólido contra un sesgo sustantivo maya más general. Sus hallazgos son consistentes con la centralidad del aprendizaje de verbos en la investigación sobre la adquisición de la lengua maya; En una variedad de lenguas mayas, los niños usan

² En particular, esta interpretación depende de tratar los nombres propios como una categoría separada. Gentner (1982; 2006) sostiene que los nombres propios son fundamentales para el análisis de un sesgo sustantivo temprano porque probablemente incluyen los objetos animados más destacados en los entornos tempranos de los niños. Si se incluyen nombres propios como sustantivos, los dos niños en el estudio de Brown (1998) *hacer* Inicialmente muestran un sesgo sustantivo, pero el inventario sustantivo es rápidamente superado por los verbos.

consistentemente verbos en sus primeras producciones y rápidamente muestran variación en el uso y la inflexión de los verbos (p. ej., Brown et al., 2013; Mateo Pedro, 2015; Pye 1985; Pye et al., 2017; Pfeiler et al., 2003). En otras palabras, el aprendizaje de verbos es el lugar obvio para examinar el desarrollo léxico y morfosintáctico temprano en las lenguas mayas.

Sin embargo, estos hallazgos provienen exclusivamente de exámenes del habla espontánea, mientras que el sesgo sustantivo se ha mostrado más consistentemente en análisis de datos de listas de verificación de vocabulario (ver, por ejemplo, Gentner y Boroditsky, 2001). Un estudio ha intentado abordar esto: Brown, Gentner y Braun (2005) elaboraron una lista de verificación de vocabulario preliminar de 594 palabras para tseltal, basada en términos generales en el CDI (283 sustantivos, 207 verbos, 104 otras palabras; ver Gentner, 2006 para una descripción completa). Luego recopilaron informes de los cuidadores de los vocabularios de cinco niños pequeños y encontraron una presencia uniformemente mayor de sustantivos que de verbos tanto en la proporción de sustantivos (media = 0,57, rango = 0,54–0,62) como en la proporción sustantivo-verbo (media = 1,34, rango = 1,17– 1.63). En particular, las palabras para seres animados constituían casi un tercio de los sustantivos. De acuerdo con el estudio de Brown (1998) sobre el habla espontánea, también encontraron una proporción relativamente alta de verbos, lo que sugiere que, incluso con evidencia de un sesgo sustantivo, los niños tseltales todavía mostraban un rápido aprendizaje de los verbos en comparación con los niños de otras lenguas estudiadas.

Si bien estos datos representan un esfuerzo verdaderamente notable para recopilar datos de listas de verificación de vocabulario para tseltal, el tamaño de la muestra aún es bastante pequeño (*norte* = 5), el cuestionario no fue validado y los investigadores encontraron desafíos tanto en la selección de formas de palabras específicas como en la obtención de respuestas seguras de los cuidadores (por ejemplo, qué inflexiones verbales usar, cuánto contexto proporcionar, cómo administrar la encuesta). Los métodos por sí solos no pueden explicar fácilmente la discrepancia entre los hallazgos de esta lista de verificación y los hallazgos del habla espontánea de Brown (1998) o de León (1999a, 1999b, 2001). En conjunto, los datos sobre el sesgo sustantivo tseltal son algo contradictorios: la evidencia en contra del sesgo sustantivo proviene de análisis del habla espontánea y la evidencia a favor proviene de los datos de la lista de verificación, mientras que la evidencia del aprendizaje temprano de verbos facilitado parece consistente en todos estos conjuntos de datos.

1.3 Representación relativa con datos de transcripción

Trabajos anteriores sobre el sesgo sustantivo en lenguas mayas se han centrado principalmente en datos del habla espontánea y, por lo tanto, han utilizado medidas de frecuencia proporcional (por ejemplo, la proporción de tipos de sustantivos y la relación sustantivo-verbo). Lo ideal sería analizar la subrepresentación y la sobrerrepresentación, pero ese enfoque sólo se ha adoptado con datos de listas de verificación. Este enfoque de representación relativa es ideal porque es el único que define claramente cómo sería el desarrollo imparcial antes de verificar la presencia de un sesgo de aprendizaje (Bates et al., 1994; 1995; Braginsky et al., 2019; Frank et al., 2021). Es importante destacar que esta elección de análisis puede tener un impacto enorme en lo que aprendemos de nuestro conjunto de datos. Considere, por ejemplo, un niño del que sus padres informan que dice 10 sustantivos (de una lista de 100) y 5 verbos (de una lista de 50): el niño está al 10% en la adquisición de cada clase de palabra (es decir, representaciones relativas iguales), pero el análisis proporcional muestra un predominio sustantivo muy fuerte (una proporción de 2:1, es decir, un fuerte sesgo nominal). La discrepancia proviene de la diferencia en el tamaño del inventario de referencia, porque simplemente hay muchos más sustantivos que aprender que verbos.

La lógica del análisis de representación relativa es algo así como lo siguiente: si asumimos que los niños adquieren un núcleo de vocabulario a cierta edad (por ejemplo, el representado por el CDI a los 30 meses), el aprendizaje imparcial parecería una adquisición uniforme e incremental dentro de un período determinado. y entre clases de palabras que alcanza el 100% al final del rango de edad representado. Por ejemplo, el paso del 0 % al 10 % se produciría en paralelo en todas las clases de palabras, independientemente de si el 10 % representa 100 palabras o 50 para una clase determinada. Eso sería un aprendizaje imparcial; por lo tanto, el aprendizaje sesgado sería siempre que el aprendizaje se desvíe significativamente de esta trayectoria imparcial. Es decir, el aprendizaje sesgado ocurriría cuando los niños saben proporcionalmente menos palabras en una clase en relación con su tamaño total de vocabulario (es decir, subrepresentación) o proporcionalmente más (es decir, sobrerepresentación).

Desafortunadamente, aún no existe una lista de verificación de vocabulario validada para tseltal (ver Brown et al., (2005) para una versión preliminar y Authors et al., (en preparación) para una lista de verificación en desarrollo), pero sostenemos que un análisis similar de vocabulario relativo La representación se puede realizar con una gran muestra de datos de transcripción. Para aproximarnos conceptualmente al análisis de representación relativa con datos de transcripción, primero debemos establecer un vocabulario básico que esperamos que hayan adquirido los niños de cierta edad. Luego debemos observar cuáles de esas palabras usan los niños dentro de cada categoría, en un rango de edad que va desde antes de que se adquieran las primeras palabras hasta la última edad representada. En lugar de una lista máxima de vocabulario (por ejemplo, el diccionario tseltal multi dialéctico; Polian, 2020), necesitamos algo parecido a un vocabulario central: una lista de palabras conocidas por casi todos los hablantes de tseltal del geolecto estudiado.

La conversación conversacional en la vida cotidiana es una fuente razonable de vocabulario conocido por la mayoría de los miembros de la comunidad. Es el nicho ecológico central para las lenguas (a diferencia, por ejemplo, de la alfabetización): es el único modo universal reconocido de uso del lenguaje y proporciona el contexto básico para el desarrollo del lenguaje infantil (Levinson, 2019). El diálogo conversacional también depende de puntos comunes comunitarios, información a la que los miembros de la comunidad pueden asumir mutuamente que los demás tienen acceso (incluido el vocabulario y las experiencias comunes; H. H. Clark 1996; Capítulo 4). Aún mejor, se produjo una conversación conversacional.*para y por niños* da una impresión aún más cercana de las palabras que se espera que los niños sepan. Por lo tanto, para estimar un vocabulario básico tenemos varias opciones: palabras utilizadas en el habla producida por niños, habla dirigida a niños o ambas. En lo que sigue utilizamos el habla producida por niños porque es nuestra mejor evidencia directa de qué vocabulario podemos esperar que utilicen en las edades observadas. Dicho esto, no estamos seguros de cuál de estas opciones es conceptualmente óptima para los propósitos actuales, por lo que también proporcionamos análisis (con resultados cualitativamente idénticos) basados en datos producidos por niños objetivo y discurso dirigido al niño objetivo en los Materiales complementarios.

Definido de esta manera, obtenemos un marco analítico con valor conceptual similar al establecido por Bates, Braginsky y Frank en sus análisis de datos de listas de verificación de vocabulario (Bates et al., 1994; 1995; Braginsky et al., 2019; Frank et al. al., 2021): creamos una lista de palabras que se espera que los niños digan en la vida cotidiana a los 52 meses de edad, y luego examinamos qué palabras realmente dicen, en un rango de edad que abarca de 9 a 52 meses. Esperamos que muchas de estas palabras estén en un equivalente tseltal de un CDI, y que los niños aprendan muchas de ellas a la edad más avanzada que observamos. Dicho esto, los datos de transcripción son muy escasos por naturaleza y los niños sólo demostrarán una pequeña fracción de las palabras que saben en los 45 minutos de grabación muestreados, por lo que no esperamos que el uso de palabras dentro de la categoría alcance

el 100%, que es lo que esperaríamos. esperar de un cuestionario estandarizado por edades como el CDI. Lo que podemos hacer en cambio es examinar la sobre y la subrepresentación en el espacio de desarrollo que podemos observar, donde los principios básicos de la sobre y la subrepresentación aún se mantienen: el aprendizaje imparcial seguiría apareciendo como un crecimiento relativo continuo a lo largo de todo el mundo. clases de palabras.

Es importante destacar que todavía no sabemos cuántos datos se necesitan para que dicho análisis sea sólido. Cuanto mayor sea el conjunto de datos y más denso el muestreo por niño, es más probable que los datos de las transcripciones se acerquen a patrones similares a listas de verificación en el análisis (Gentner, 1982; 2006 presenta un argumento similar con respecto al “vocabulario acumulativo”). El presente estudio es un intento preliminar de utilizar este enfoque analítico con la mayor colección de transcripciones existente del habla tseltal cotidiana (Casillas, 2023). Combinamos este enfoque tan novedoso con uno más tradicional (la frecuencia proporcional) para examinar de manera integral la evidencia de un sesgo sustantivo en el desarrollo del vocabulario tseltal.

1.4 El estudio actual

Analizamos el uso de raíces nominales y verbales en el habla espontánea de 29 niños tseltales entre 0;9 y 4;4 años durante el transcurso de días enteros de vigilia en casa. Primero examinamos la frecuencia proporcional de sus usos de sustantivos y verbos para establecer comparaciones directas con estas mismas estimaciones de trabajos anteriores. Luego volvemos a examinar los mismos datos utilizando un enfoque de representación relativa, para el cual no tenemos una comparación previa directa. Los datos de transcripción utilizados aquí provienen de un estudio más amplio y en curso sobre el desarrollo temprano del lenguaje en tseltal (Casillas, 2023). Con base en el trabajo resumido anteriormente, predijimos un sesgo sustantivo atenuado en tseltal con evidencia de aprendizaje temprano de verbos.

2. Métodos

2.1 Participantes y comunidad de estudio

Estos datos provienen de una colección de grabaciones realizadas en 2015 de niños mayas tseltales menores de 5 años (Casillas et al., 2017a). La colección contiene segmentos transcritos de una muestra transversal de 41 grabaciones de niños de 4; 4 años y menos (rango = 2 a 52 meses; media = 21,56 meses; mediana = 19 meses). En los análisis solo usaremos datos de las grabaciones de 29 niños de 9 meses de edad y mayores (consulte la Figura 2 para ver los puntos de datos que muestran las edades individuales). Todos los niños de esta muestra provienen de un solo pueblo rural del sur de Chiapas. En esta comunidad, el tseltal es el idioma principal que se habla (geolecto de Tenejapan) y la agricultura de subsistencia es la ocupación principal de la mayoría de los residentes adultos. Dicho esto, muchas familias tienen un adulto que obtiene ingresos por otros medios (por ejemplo, conduciendo un taxi, trabajando fuera de la comunidad, etc.). Todos los niños de la muestra crecían hablando tseltal como lengua materna principal, aunque la mayoría también escuchaba hablar español con regularidad en público (por ejemplo, cerca de la escuela), en algunos discursos de adultos que se pueden escuchar, en programas de radio y televisión, y en el préstamo de palabras españolas al tseltal (p. ej., *con/auto* para auto). Los niños de esta muestra eran demasiado pequeños para asistir a los servicios escolares y, por lo tanto, normalmente pasaban sus días en compañía de miembros de la familia, a menudo en casa (pero no siempre;

por ejemplo, pasando horas en el huerto familiar, visitando casas de otras personas, haciendo un recado). , etc.).

En esta comunidad, los bebés menores de un año pasan la mayor parte del día en un portabebés que lleva su madre o, a veces, otra pariente cercana, donde se les puede reposicionar rápida y fácilmente para dormir (de espaldas), alimentarse (de frente) y observar el movimientos a su alrededor (cadera/espalda) mientras la persona que los lleva se dedica a sus tareas diarias. Más adelante en el desarrollo, entre los 3 y los 5 años, los niños comienzan a pasar mucho más tiempo en compañía de otros niños, generalmente hermanos, primos y otros vecinos. La mayor parte de la vivienda está organizada patrilocalmente, de modo que los niños crecen cerca de los padres de su padre y de las familias de sus hermanos. En nuestra base de datos, el intervalo típico entre niños es de alrededor de 3 años y el tamaño típico del hogar es de 7 personas, incluido el niño objetivo (los miembros del hogar se definen aquí como el número de personas que comparten una cocina u otro espacio vital principal; rango = 3-15). Por esta razón, podemos esperar que la experiencia lingüística diaria típica de un bebé en el conjunto de datos sea bastante diferente a la de un niño mayor, lo que se refleja en las fluctuaciones de la tasa de entrada a lo largo del día (Casillas et al., 2020). En resumen, el tsetal es la principal fuente de información lingüística de estos niños; esta información proviene principalmente de miembros de la familia y, cada vez más con la edad, proviene de otros niños.

2.2 Procedimiento de registro

Los participantes fueron reclutados de boca en boca durante aproximadamente cuatro semanas por un miembro local de la comunidad que trabajó con [AUTOR]. El consentimiento se realizó de forma verbal e iterativa, exclusivamente en tsetal: en el momento del reclutamiento inicial, el asistente de investigación local describió la tarea de grabación y el plan de compensación, incluido lo que se captura en el dispositivo de grabación y cualquier información sobre el investigador visitante ([AUTOR]) que el participante tenía curiosidad; luego, antes del día de grabación planificado, el asistente de investigación recordó a la familia participante sobre la tarea y lo que se capturaría en el dispositivo de grabación; se notificó a los familiares que no querían ser grabados para que pudieran ir a otro lugar el día de la grabación; Antes del inicio de la grabación, se llevó a cabo un proceso de consentimiento más formalizado, repasando la tarea, la compensación, los derechos de los participantes y el uso futuro de los datos de la grabación. Los participantes pudieron hacer preguntas y comprobar su comprensión en cada punto del consentimiento. El día de la grabación, el asistente de investigación y [AUTOR] visitaron a la familia participante para recopilar información demográfica sobre el niño objetivo y su familia mediante una entrevista antes de iniciar la grabación.

Una vez que se dio el consentimiento oral (que normalmente no se registra), el asistente de investigación se acercó al niño objetivo y le colocó un arnés que incluía una grabadora de audio ya en vivo (Olympus WS-832) y una cámara de pecho que tomaba fotografías cada 30 segundos (Narrative Clip 1 + lente Photojojo Super Fisheye). Luego, el asistente de investigación demostró cómo usar la solapa de privacidad del arnés, una cubierta para la lente que se podía colocar o quitar rápidamente en caso de que los participantes quisieran bloquear temporalmente la visión de la cámara. Se alentó a los cuidadores a usar la cubierta si querían privacidad, pero en caso contrario intentarían dejar la lente descubierta. Se pidió a los cuidadores que le quitaran temporalmente el arnés al niño si perturbaba su sueño o si el niño iba a mojarse excesivamente y, en esos casos, simplemente lo colocaran cerca del niño. La intención de una grabación de un día de duración es obtener una estimación del lenguaje típico de los niños en el hogar. En realidad, el día de la grabación sigue siendo un especial día en el que ciertos miembros de la familia pudieron haber elegido no estar presentes y durante el cual el arnés sirvió como un recordatorio visual continuo de que se estaban documentando la vista

frontal y el entorno auditivo proximal del niño. La grabadora y la cámara fueron comentadas con frecuencia durante el día de grabación en muchas grabaciones de niños.

En esta recopilación de grabaciones de 2015, se realizaron un total de 57 grabaciones de 55 niños nacidos de 43 madres (se realizaron grabaciones de recuperación en caso de falla de la cámara/grabadora de audio en la primera grabación). Las grabaciones generalmente duraban desde media mañana hasta el atardecer, con una duración media de 8,96 horas de datos de grabación de audio (rango = 3,0 a 10,2) y una hora de inicio típica de 08:30 (rango = 07:18 a 10:02). Las transmisiones de audio y fotografías sin procesar se alinearon manualmente y se combinaron digitalmente en archivos de video MP4 en 2016.

2.3 Muestreo y anotación de datos

Siguiendo a Casillas y colegas (2020; 2021), examinamos el habla espontánea producida en nueve clips de 5 minutos que fueron seleccionados al azar del lapso de cada grabación (es decir, 45 minutos en total por grabación). Los clips no se superponen y se seleccionaron independientemente de la actividad en curso (por ejemplo, el niño podría estar durmiendo durante un clip) para obtener una vista totalmente representativa de la actividad del lenguaje durante las horas del día.

Casillas y colegas (2020) transcribieron sus clips utilizando el esquema de anotación ACLEW, que anota toda voz audible en la grabación de audio con indicadores de madurez vocal del destinatario y del niño (Casillas et al., 2017b; Bunce y otros, 2020; Soderstrom et al., 2021). Nuestro objetivo aquí era obtener una muestra mucho más grande de datos de lenguaje, por lo que creamos una adaptación mínima de su esquema de anotación que incluye transcripciones de todas las vocalizaciones producidas por el niño objetivo (CHI) y todas las vocalizaciones dirigidas por el niño objetivo (OTR) que aparecen en los clips. Una diferencia principal en el flujo de trabajo de anotación es que los datos presentes se limitan exclusivamente a entradas dirigidas al niño objetivo.

Trabajos anteriores muestran que estos niños tseltales encuentran mucho más input lingüístico más allá de lo que podemos analizar aquí: el input dirigido por otros (por ejemplo, el habla entre adultos o dirigido a otros niños) supera el input dirigido por el niño objetivo en una proporción de seis a uno (Casillas et al., 2020). Por lo tanto, nuestra elección de transcribir únicamente información dirigida al niño objetivo limita cualquier análisis a sólo una porción de lo que los niños pueden aprender en sus entornos lingüísticos, una limitación que podría tener más impacto para los niños mayores que para los más pequeños (por ejemplo, Akhtar, 2005; Arunachalam, 2013). Sin embargo, como argumentamos anteriormente, el habla producida y dirigida por el niño objetivo son las fuentes de vocabulario más probablemente relevantes para los presentes análisis. Esta misma decisión también nos permite examinar de manera factible los datos de más niños. La notación y transcripción completa de todo el habla escuchable para grabaciones de un día en tseltal requiere dos anotadores (un hablante nativo y un experto en ELAN) y requiere aproximadamente 50 minutos de trabajo por cada minuto de audio grabado (Casillas et al., 2020). Al reducir nuestro objetivo de transcripción y anotación a discursos dirigidos y producidos por niños, la tarea se acelera enormemente y se vuelve posible durante todo el año como un proyecto dirigido por un solo hablante nativo de tseltal. Aun así, los datos de transcripción analizados aquí representan más de dos años de trabajo enfocado.

Para los fines actuales, las anotaciones de las expresiones transcritas se complementaron de la siguiente manera: traducciones sueltas de cada expresión al español (cucharadita), una indicación del tipo de hablante para todo el discurso dirigido al niño objetivo (quien; ej., mujer/hombre/niña/niño/otro) y una lista de cada raíz de sustantivo (sus) y verbo (ver) utilizada en cada vocalización (ver Figura 1).

Toda la transcripción, traducción y anotación de sustantivos y verbos fue completada por [AUTOR], un hablante nativo de Tenejapan Tseltal que reside fuera del pueblo estudiado. Toda la transcripción y anotación se realizó en el software de anotación de video ELAN (Wittenburg et al., 2006). De las 57 grabaciones originales de la colección de 2015, aquí incluimos los 29 a los que se aplica este canal de anotaciones y para los cuales el niño objetivo tiene 9 meses o más (rango = 9 a 52 meses; media = 28.44; mediana = 29; desviación estándar = 12,86). Utilizamos 9 meses como punto de partida porque es probable que capture de manera conservadora el inicio de la producción de la primera palabra, que normalmente esperaríamos que ocurriera alrededor de los 12 meses (Casillas et al., 2020; Schneider et al., 2015).. Hacemos *no* analice las 10 grabaciones adicionales procesadas utilizando el esquema de anotación ACLEW por Casillas y colegas (2020) dadas las diferencias significativas entre su flujo de trabajo y el equipo de anotación y el nuestro ([AUTOR] y un asistente de investigación local que conocía a todas las familias).

Si bien creemos que las transcripciones y anotaciones son en general de muy alta calidad, observamos que este corpus es un recurso vivo que sirve para múltiples proyectos de investigación simultáneamente. Esperamos agregar gradualmente anotaciones para otros tipos de entrada del lenguaje, otras características lingüísticas del habla producida y pequeñas correcciones a la segmentación y transcripción en las próximas décadas. Por lo tanto, los datos contenidos en este artículo deben tomarse como una instantánea de un recurso vivo y en crecimiento. En última instancia, planeamos replicar y ampliar estos análisis con la grabación de audio completa (*norte* = 57) conjunto de datos y con una lista de verificación de vocabulario tseltal validada.

Level	Transcription	Translation
CHI [33]	kayben	kayben i
tisp@CHI [27]	abremelo	abremelo
ver@CHI [33]	kay (abrir)	kay (abrir)
sus@CHI [33]	NO-HAY-SUSTANTIVOS	NO-HAY-SUSTANTIVOS
comentarios [7]		
OTR [11]	ya jlok'taat tali ya x-ok'ati, ok'an kiltik i	ja' lek uch'a atole@s:s
quien@OTR	mujer	mujer
tisp@OTR [11]	te voy a grabar así como estas llorando, llora pues	mejor toma atole
ver@OTR [11]	lok'ta (grabar), ok' (llorar)	uch' (tomar)
sus@OTR [11]	NO-HAY-SUSTANTIVOS	atole (atole)

Figura 1. Captura de pantalla de aproximadamente 5 segundos de un archivo de transcripción que muestra la estructura jerárquica de niveles utilizada para organizar la transcripción, la traducción y la anotación de sustantivos y verbos. Los medios vinculados para cada clip incluían un archivo de audio.wav y un archivo de video.mp4 que mostraba, cuando era posible, fotografías de cámaras de pecho usadas por niños alineadas en el tiempo.

Locución infantil completa y glosa.	Sustantivos	Verbos
stsek alal s-tsek alal 3POSS-falda bebe 'la falda del bebé'	tsek (falda cruzada), alal (bebé)	NA

<i>jtsaktik jilel ku'untik eki</i> <i>j-tsak-tik jilel k-u'un-tik ek=yo</i> 1ERG-take-PL DIR 1POSS-REL.N-PL también=PARTE 'nosotros también conservamos el nuestro'	NA	tsak (tomar)
<i>uk'um jpastik</i> <i>uk'um j-pas-tik</i> río 1ERG-do-PL 'lo hacemos (en) el río'	uk'um (río)	no (hacer/hacer)

Tabla 1. Ejemplos de tres expresiones producidas por niños en el conjunto de datos, incluida una glosa para cada una y los sustantivos y verbos extraídos de cada una para el presente análisis. Abreviaturas glosadas como sigue: DIR: direccional; ERG: ergativo; PARTE: partícula; PL: plural; POSS: posesivo; REL.N: sustantivo relacional.

2.4 Preparación de datos

Para preparar las transcripciones para el análisis, utilizamos la biblioteca `phonfieldwork` en R (Moroz, 2021) para convertir los datos transcritos en archivos `.eaf` a un formato tabular. Luego creamos resúmenes tabulares del uso de sustantivos y verbos utilizando la biblioteca `tidyverse` (Wickham et al., 2019). Luego, [AUTOR] exportó las raíces de verbos y sustantivos y las anotó manualmente para que coincidieran con las formas estandarizadas y los metadatos del diccionario multidialectal abierto en línea `tseltal-español` (Polian, 2020). Este paso de anotación manual es necesario dada la falta actual de un programa de derivación automática para `tseltal` y la variabilidad en las formas ortográficas. En este proceso de anotación manual, [AUTOR] también proporcionó metadatos adicionales sobre la clasificación de los verbos para informar los análisis finales. Se consideró que los sustantivos incluían todas las instancias de: sustantivo, sustantivo de acción, sustantivo compuesto y sustantivo propio. Tenga en cuenta que los términos de parentesco no se clasificaron por separado cuando se usaron vocativamente (es decir, cuando se usaron como nombres propios, “¡Mami, ven!” vs. “Esa es mi mami”). Aquí tratamos todos los usos de términos de parentesco como instancias de sustantivos básicos. Se consideró que los verbos incluían todas las instancias de: verbos intransitivos agentes, verbos intransitivos defectuosos, verbos intransitivos de movimiento, verbos intransitivos fasales, verbos intransitivos, verbos transitivos defectuosos, verbos transitivos y auxiliares. Estas clases de palabras estandarizadas se derivaron del diccionario multidialectal abierto en línea `tseltal-español` (Polian, 2020). Estos datos verificados manualmente se volvieron a leer en R para su análisis estadístico. Nuestros análisis de sobrerrepresentación y subrepresentación se basan en el código R disponible abiertamente en el libro `WordBank` (Braginsky et al., 2019; Frank et al., 2021). Todas las parcelas se crean con el `ggplot` La biblioteca (Wickham, 2016) y los análisis estadísticos se ejecutan con la biblioteca `lme4` (Bates et al., 2015). Se creó una versión reproducible de este manuscrito utilizando `elpapaya` y bibliotecas `Rmarkdown` en la aplicación de escritorio `Rstudio` (Aust & Barth, 2022), que está (estará) disponible con datos anonimizados en un repositorio de Open Science Foundation (https://osf.io/9gmzr/?view_only=eff1a29f8f9240159640fb21c966f5a9).

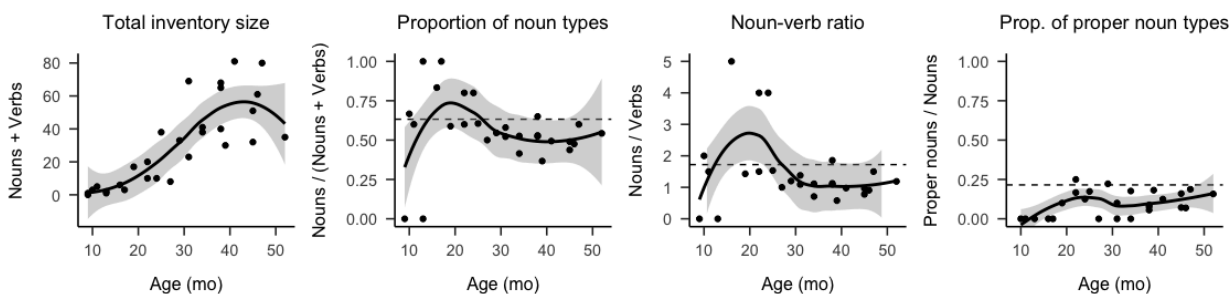


Figura 2. Medidas de inventario sin procesar de raíces de sustantivos y verbos producidas por niños, trazadas por edad del niño, que incluyen: tamaño total del inventario de raíces de sustantivo-verbo (izquierda), proporción de raíces únicas que son sustantivos (centro izquierda), proporción de raíces de sustantivos únicas a raíces verbales únicas (centro-derecha) y proporción de raíces nominales únicas que son nombres propios (derecha). Las estimaciones de cada medida del conjunto total de raíces verbales y nominales únicas observadas (es decir, en todos los discursos producidos por el niño objetivo y en todos los discursos dirigidos por el niño objetivo) se muestran como líneas horizontales discontinuas, excepto el tamaño total del inventario (NB: el total El inventario agrupado es de 546 raíces de sustantivos y verbos únicos). Las líneas negras gruesas indican un ajuste suavizado utilizando la función `loess()` de la `ggplot` biblioteca en R (Wickham, 2016) para las cuales las regiones sombreadas indican intervalos de confianza estimados del 95%.

Estudiar	Características de la muestra	Proporción de tipos de sustantivos $\frac{\#Tipos_N}{\#Tipos_N + \#Tipos_V}$	Relación sustantivo-verbo $\frac{\#Tipos_N}{\#Tipos_V}$	Proporción de tipos de nombres propios $\frac{\#Tipos_{PropN}}{\#Tipos_N}$	Notas
Este estudio	<ul style="list-style-type: none"> Idioma: Tenejapa tseltal norte = 29 (9–52 meses) Fuente: Grabaciones de audio transversales de un día; 9 clips de 5 minutos muestreados aleatoriamente 	0,56	1,48	0.096	<ul style="list-style-type: none"> Basado en niños que se observó que producían al menos un sustantivo o verbo
Caballero (1982)	Idioma: chino mandarín, japonés, kaluli, inglés, alemán y turco norte = 16 (norte = 2–4 por idioma; 14 a 29 meses) Fuente: Transcripciones glosadas en inglés de grabaciones caseras (kaluli, chino mandarín, turco) o vocabulario acumulativo obtenido por el cuidador (todos los demás; consulte Gentner (1982) Apéndice A para más detalles)	0,68–0,85* Kaluli: 0,68 Chino mandarín: 0,70 Turco: 0,74 Japonés: 0,79 Inglés: 0,82 Alemán: 0,85	2,3–6,17* Kaluli: 2.30 Chino mandarín: 2,31 Alemán: 2,44 Turco: 3,10 Japonés: 4,33 Inglés: 6.17	0,15–0,88 Inglés: 0,15 Alemán: 0,18 Turco: 0,29 Japonés: 0,36 Chino mandarín: 0,41 Kaluli: 0,88	<ul style="list-style-type: none"> * Basado en la Tabla 4; en lugar de tipos de verbos únicos, estos cálculos utilizan tipos de predicados únicos, que incluyen verbos, preposiciones y modificadores. ** Basado en la Tabla 5; solo incluye datos de las primeras palabras de un niño por idioma

Marrón, marrón y caballero (2005)	Idioma: Tenejapa tseltal <i>norte</i> = 5 ("niños pequeños"; edad no informada) Fuente: Informe del cuidador en una lista de verificación de vocabulario tseltal con 594 palabras (283 sustantivos, 207 verbos, 104 otras palabras)	0,57***	1.34	0,30	<ul style="list-style-type: none"> *** Esta proporción se calcula sobre el tamaño total del vocabulario (no solo Ns+Vs) y, por lo tanto, es probable que esté subestimada en relación con los demás valores de su columna.
Marrón (1998)	<ul style="list-style-type: none"> Idioma: Tenejapa tseltal <i>norte</i> = 2 (15 a 29 meses) Fuente: Grabaciones de audio domésticas longitudinales mensuales con grabaciones de vídeo domésticas ocasionales. 	0,53	1.17	0,16	<ul style="list-style-type: none"> Basado en la Tabla 1, tratando las N apropiadas como N y excluyendo "Otros" Estimaciones promediadas entre las etapas de MLU y los niños
de León (1999b)	<ul style="list-style-type: none"> Language: Zinacantán Sobre Tsotsi <i>norte</i> = 2 (20–25 meses) Fuente: Grabaciones longitudinales de audio o video en el hogar, semanales o quincenales, de visitas de familiares extendidos durante el período de la muestra. 	0,46	0.85	0,12	<ul style="list-style-type: none"> Con base en la Tabla 3, se tratan los N apropiados y los términos de parentesco (reportados conjuntamente con los N apropiados) como N y se excluyen todas las demás categorías.

Tabla 2. Resumen de las estimaciones de frecuencia proporcional reportadas en este estudio y en estudios previos sobre el desarrollo de sustantivos y verbos mayas (Brown, 1998; de León, 1999b). Las dos medidas comparables incluyen: (A) Proporción de tipos de sustantivos (aquí definidos como el número de tipos de sustantivos únicos dividido por la suma del número de tipos de sustantivos únicos y el número de tipos de verbos únicos); (B) la relación sustantivo-verbo (aquí definida como el número de tipos de sustantivos únicos dividido por el número de tipos de verbos únicos); y (C) la proporción de tipos de nombres propios (definida aquí como el número de tipos de nombres propios únicos y términos de parentesco dividido por el número de todos los tipos de nombres únicos). En (A) y (B), los "tipos de sustantivos" incluyen nombres propios y términos de parentesco, aunque observamos que este es un punto importante de discordia entre los estudios citados. En las ecuaciones de ejemplo anteriores, "N" = todos los sustantivos (incluidos los nombres propios y términos de parentesco), "V" = todos los verbos y "PropN" = sustantivos subclasificados como nombres propios y términos de parentesco.

3. Resultados

A continuación, primero daremos una descripción general de los datos de sustantivos y verbos disponibles en el corpus. Luego revisamos los hallazgos dado el enfoque tradicional de frecuencia proporcional, haciendo comparaciones directas con trabajos anteriores. Después de

** Este manuscrito fue traducido del inglés con la ayuda de [Google Translate](https://www.google.com/translate).

eso, revisamos los hallazgos nuevamente, esta vez utilizando el enfoque de representación relativa más reciente, que se adapta aquí para los datos de transcripción.

El conjunto de datos incluye 11485 vocalizaciones: 9011 producido por el niño objetivo, 1677 producido por un adulto, y 797 producido por otro niño. De estos, 2912 los enunciados contenían al menos un sustantivo o un verbo: 1811 de las expresiones del niño objetivo (20.1%), 784 de las expresiones adultas (46.7%), y 317 de las expresiones de los otros niños (39.8%; consulte los Materiales complementarios para obtener un análisis utilizando estas raíces del habla dirigida al niño objetivo).

Entre el 29 niños objetivo, 28 produjo al menos un sustantivo o verbo en los 45 minutos observados; Se observó que todos los niños mayores de un año producían un sustantivo o un verbo (rango: 1-170). La edad más temprana a la que se registró un sustantivo producido por un niño fue de 10 meses (*Chen's* 'frijol' y mamá 'mamá') y la edad más temprana a la que se registró un verbo producido por un niño fue de 9 meses (*'Agarrar'*; y 'En segundo lugar quedó 'dar/poner', con una producción de 10 meses). Estas primeras raíces nominales y verbales concuerdan con trabajos anteriores que postulan que los términos de parentesco se encuentran entre los muy primeras palabras observadas en la producción (por ejemplo, Brown 1998; de León, 1999a,b; 2001; Genner 1982; 2006; Schneider y otros, 2015) mientras que estos primeros usos de verbos "ligeros" generales contrastan con los hallazgos anteriores de Brown (1998) sobre tseltal.

3.1 Medidas de frecuencia proporcional

Primero examinamos los valores de frecuencia proporcional en el conjunto de datos actual, luego los comparamos con trabajos anteriores sobre lenguas mayas y luego con aquellos de trabajos anteriores sobre otras lenguas. Estimaciones por niño del uso de sustantivos y raíces verbales en el habla producida por niños se resumen en la Figura 2, y los valores promedio de estos datos y conjuntos de datos anteriores comparables se resumen en la Tabla 2.

Colectivamente los niños produjeron 248 tipos de sustantivos y 138 tipos de verbos. El total de raíces de sustantivos y verbos (es decir, el tamaño del vocabulario) observado para cada niño aumentó con la edad (Figura 2, panel más a la izquierda). La proporción de tipos de sustantivos en los inventarios de niños individuales promedió 0.56 (rango: 0,0–1,0; dos niños con inventarios muy pequeños produjeron solo verbos; Figura 2, panel central izquierdo) y la proporción de sustantivos a verbos promedió 1.48 (rango: 0–5; dos niños con inventarios muy pequeños produjeron solo sustantivos; Figura 2, panel central derecho). Si consideramos que un sesgo de sustantivo es cualquier proporción de sustantivos superior a 0,5, 19 de los 28 niños que produjeron uno o más sustantivos o verbos mostraron un sesgo sustantivo (67,9%). Si consideramos que un sesgo de sustantivo es cualquier relación sustantivo-verbo superior a 1,0, nuevamente 19 de los 28 niños que produjeron uno o más sustantivos o verbos mostraron un sesgo sustantivo (67,9%).

Se predice que el sesgo sustantivo será más evidente en el período más temprano de producción de palabras (especialmente dentro de las primeras 50 a 100 palabras; Caballero, 2006), que, según los datos multi lingüísticos, normalmente se puede alcanzar entre 16 y 20 meses (Frank et al., 2021: Capítulo 5.1). Por lo tanto llevamos a cabo *t*-pruebas de nuestros datos que comparan la proporción de tipos de sustantivos y la proporción de sustantivos-verbos antes y después de los 20 meses de edad. No encontramos evidencia de una diferencia entre las medidas de sesgo de sustantivo temprano (≤ 20 meses) y posterior (más de 21 meses) entre los niños que produjeron al menos un sustantivo o verbo (proporción de tipos de sustantivo: $t(7,429) = 0,253$, $p = 0,807$, *IC del 95%* [-0,296, 0,368]; relación sustantivo-verbo: $t(5,79) = 0,296$, $p = 0,778$, *IC del 95%* [-1.690, 2.151]). Un examen detenido de la Figura 2 hace evidente que, en las edades más tempranas documentadas, los niños

exhibieron una variación dramática en su uso proporcional de sustantivos, con dos niños usando categóricamente sustantivos y dos niños usando categóricamente verbos.

¿Cómo se comparan estos hallazgos con trabajos anteriores? Dejando a un lado las diferencias de edad y preprocesamiento de datos, los resultados son comparables (si no ligeramente más favorables al sesgo sustantivo) en comparación con trabajos anteriores sobre lenguas mayas (Brown, 1998; Brown et al., 2005; de León, 1999b; Tabla 2). En trabajos que utilizan el habla espontánea, los investigadores de la lengua maya han preferido contar los nombres propios y los términos de parentesco por separado de la categoría nominal básica (razonando de manera diferente a los estudios de Gentner (1982; 2006)), lo que en consecuencia produce proporciones nominales generales más bajas. Sin embargo, incluso si contamos esos nombres propios y términos de parentesco como sustantivos básicos (Tabla 2), nuestros resultados actuales todavía parecen más favorables a un sesgo sustantivo de lo que se informó anteriormente: los datos tseltales de Brown (1998) mostraron una proporción de tipos de sustantivos de 0,53 y una proporción de 1,17. relación sustantivo-verbo ($\Delta = -0,03$ y $-0,31$ a partir de los datos actuales) y de León (1999b) Sobre Tsotsi los datos mostraron una proporción de 0,46 y una proporción de 0,85 ($\Delta = -0,10$ y $-0,63$ de los datos actuales). Los resultados actuales son más comparables con los datos de la lista preliminar de verificación de vocabulario tseltal (Brown et al., 2005): una proporción de 0,57 de tipos de sustantivos y una proporción de sustantivo-verbo de 1,34 ($\Delta = +0,01$ y $-0,14$ de los datos actuales). aunque la proporción de tipos de sustantivos en ese estudio se calcula sobre el tamaño total del vocabulario (es decir, incluyendo una categoría "otra") y, por lo tanto, está algo subestimada en relación con nuestro cálculo actual.

También podemos comparar aproximadamente los resultados actuales con los informados en el estudio fundamental de Gentner (1982) sobre chino mandarín, japonés, kaluli, alemán, inglés y turco (*norte* = 2–4 niños por lengua). En ese artículo, los verbos se incluyen en la categoría de predicados (que incluye verbos, preposiciones y modificadores), de modo que sus resultados relacionados con los verbos parecerán más generosos que los nuestros (es decir, solo verbos). No obstante, vemos que los resultados mayas, en el presente trabajo y en estudios anteriores, son menos favorables al sesgo sustantivo que lo que se informó para estas otras seis lenguas (derivado de la Tabla 4 de Gentner (1982)): un rango translingüístico de 0,68–0,85 en proporción de tipos de sustantivos ($\Delta = +0,12$ – $0,29$ de los datos actuales) y un rango de 2,3–6,17 en la proporción de sustantivo-predicado ($\Delta = +0,82$ – $4,69$ de los datos actuales de proporción de sustantivo-verbo). Como recordatorio, los datos de Gentner (1982) utilizan una categoría de “predicado” en lugar de una categoría de “verbo”, por lo que si pudiéramos calcular la proporción y la razón sólo con verbos, probablemente mostraremos un sesgo nominal aún más fuerte en estos seis idiomas en comparación. a los conjuntos de datos mayas.

Un sub componente esencial del debate sobre el sesgo sustantivo se centra específicamente en la presencia temprana de nombres propios, por lo que también comparamos brevemente estos datos proporcionales. En conjunto, los niños de nuestro conjunto de datos produjeron 248 tipos de sustantivos, de los cuales 44 se clasificaron como nombres propios (Figura 2, panel más a la derecha). En particular, estos 44 nombres propios incluyen variaciones de forma en un subconjunto de los nombres propios observados (por ejemplo, “manuel”, “Manel” y “Man” como variaciones del nombre Manel); las 44 formas de nombres propios se basan en aproximadamente 32 nombres de personas únicos. Entre los niños que produjeron una o más raíces sustantivas, la proporción media de sustantivos que eran nombres propios fue 0,096 (rango: 0,00–0,25). Esto es inferior a lo que se ha informado en todos los estudios anteriores. Es más comparable a análisis anteriores de datos del habla espontánea de niños mayas (Tseltal (Brown, 1998): 0,16, $\Delta = +0,064$; y sobre Tsotsi (de León, 1999b): 0.12, $\Delta = +0.024$) pero menor que la estimación basada en listas de verificación para

Tseltal (Brown et al., 2005; 0.30, $\Delta = +0.204$) y menor que el rango de valores reportados para los seis idiomas examinados en Gentner (1982: 0,15–0,88, $\Delta = +0,054$ –0,694).

3.2 Representación relativa

Como se explicó anteriormente, el inconveniente de utilizar la proporción de tipos de sustantivos o la relación sustantivo-verbo como medidas primarias de un sesgo sustantivo es que no toman en cuenta la prevalencia inicial de sustantivos y verbos y, por lo tanto, no pueden hablar directamente de la prevalencia excesiva de sustantivos y verbos, o subrepresentación. Por el contrario, el enfoque más novedoso de la “representación relativa” define claramente el aprendizaje imparcial y, por lo tanto, proporciona un camino para examinar *sesgadas* trayectorias de aprendizaje. Como se explicó anteriormente, este enfoque analítico se origina en Bates (1994, 1995) y fue popularizado recientemente por Braginsky, Frank y colegas (2019; 2021), quienes lo han utilizado para confirmar una sobrerrepresentación constante de sustantivos y una subrepresentación de predicados. (incluidos los verbos) en los datos de la lista de verificación de vocabulario productivo de 27 idiomas. Si bien la representación relativa no se ha utilizado previamente con datos de transcripción, argumentamos anteriormente que puede aproximarse conceptualmente si se cuenta con suficientes datos de transcripción, con la salvedad de que no esperamos que el aprendizaje dentro de una categoría alcance cerca del 100%, incluso en las edades más avanzadas. En nuestra muestra, dada la escasez de raíces observadas en el habla espontánea.

En todo el conjunto de datos de discurso producido por el niño objetivo³, nosotros observamos 248 raíces de sustantivos únicos y 138 raíces verbales únicas, lo que nos da un tamaño total de inventario de raíces de 386. Estos 386 Las raíces constituyen nuestro vocabulario principal, con el que comparamos los usos observados de las raíces verbales y sustantivas de cada niño. Para ilustrar el proceso de puntuación, damos un ejemplo concreto de una niña de 19 meses: se observó que usaba 10 raíces nominales únicas y 7 raíces verbales únicas, por lo que el tamaño relativo total de su inventario es 0,044 (el inventario infantil total dividido por inventario básico total; $(10+7)/386$), su pariente *sustantivo* El tamaño del inventario es 0,040 (inventario total de sustantivos secundarios dividido por el inventario total de sustantivos básicos; $10/248$), y su tamaño relativo *verbo* El tamaño del inventario es 0,051 (inventario total de verbos secundarios dividido por el inventario total de verbos centrales; $7/138$). Usando este proceso, calculamos un tamaño relativo total del inventario, un tamaño relativo del inventario de sustantivos y un tamaño relativo del inventario de verbos para cada niño en el conjunto de datos.

Siguiendo a Braginsky, Frank y colegas (2019; Frank et al., 2021), para cada tipo sintáctico, ajustamos un modelo lineal restringido de representación relativa que predice la proporción de tipos adquiridos dentro de una clase de palabras. (eje y en la Figura 3) por la proporción de tipos adquiridos en todo el vocabulario (eje x en la Figura 3). Como en trabajos anteriores, nuestros modelos incluyen polinomios de tercer orden para permitir una variedad de ajustes cóncavos o convexos y estaban restringidos para predecir que los niños habrán adquirido el 100% de la categoría cuando hayan adquirido el 100% del vocabulario total.⁴—una restricción que aún se mantiene incluso si esperamos ver solo una fracción de la trayectoria de

³ En los Materiales complementarios ofrecemos una versión alternativa de este análisis en la que el vocabulario básico se basa tanto en el habla dirigida por el niño objetivo como en el discurso producido por el niño objetivo; este cambio en la operacionalización del vocabulario básico produce resultados cualitativamente similares. Atribuimos los resultados similares al bajo volumen general de discurso dirigido al niño objetivo, que solo agrega 96 raíces nominales y 61 raíces verbales a esta versión combinada del vocabulario básico (28,5% y 30,7% de las categorías, respectivamente).

⁴ $\text{clm}(\text{prop_categoría} \sim \text{l}(\text{prop_vocab}^3) + \text{l}(\text{prop_vocab}^2) + \text{vocabulario} - 1)$

desarrollo entre 0 y 100%. Como era de esperar, las raíces utilizadas observadas fueron escasas en comparación con el conjunto de vocabulario básico: el niño con el mayor vocabulario total sólo había adquirido 0.21 del inventario total observado de raíces de sustantivos y verbos. Por lo tanto, nuestro conjunto de datos completo reside en la esquina inferior izquierda del ajuste restringido de (0,0) a (1,1) y examinamos la representación insuficiente y excesiva en esta primera parte de la trayectoria de desarrollo. Observamos que el rango observado limitado sesga de manera conservadora en contra de encontrar una diferencia consistente con el aprendizaje imparcial, dado que el modelo se ajusta al rango completo entre 0 y 100% de aprendizaje.

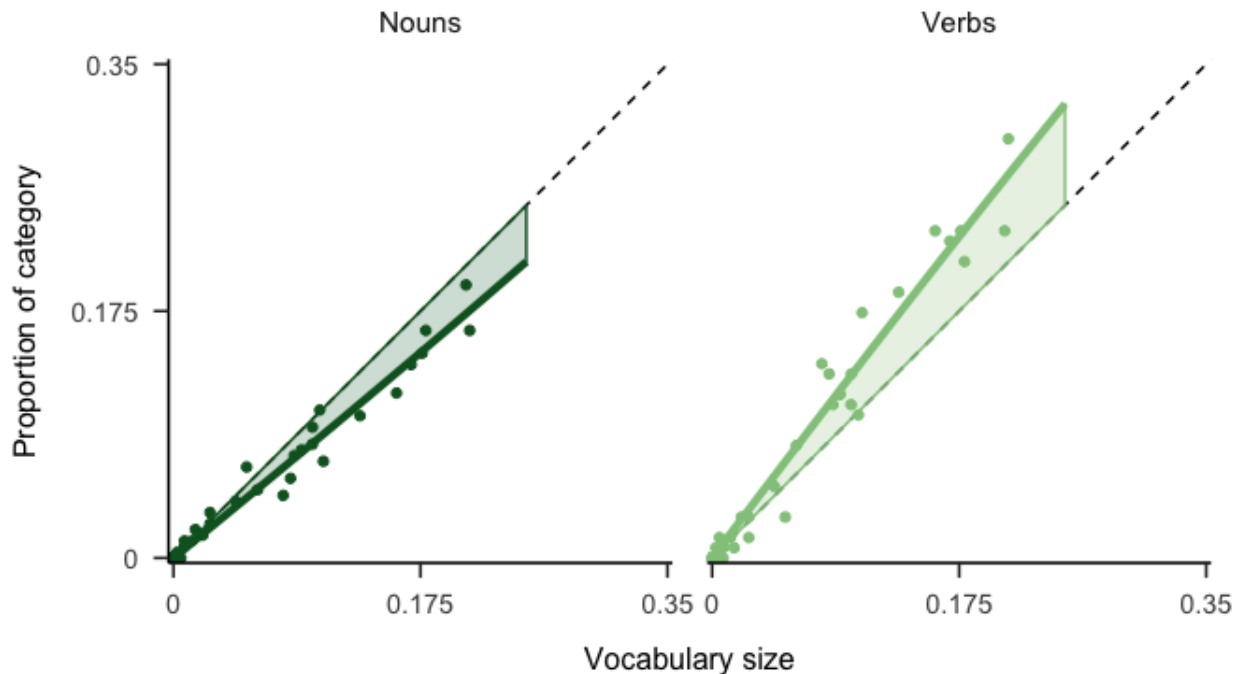


Figura 3. Representación relativa de sustantivos (izquierda) y verbos (derecha): Proporción de raíces adquiridas para una categoría dada comparada con la proporción de raíces adquiridas para el inventario total de raíces de sustantivos y verbos. Cada punto de datos representa a un niño individual, y la línea gruesa muestra el ajuste restringido del modelo, con las regiones sombreadas resaltando el área estimada entre la trayectoria de desarrollo imparcial esperada (la línea diagonal) y los datos reales observados (los puntos de datos). en el rango de tamaños de vocabulario observados (tamaño relativo total del vocabulario = [0,00:0,21]).

Dentro de los 29 niños individuales, 21 niños (72,4%) muestran una representación insuficiente de los sustantivos (es decir, debajo de la diagonal; media = -0,02) y 21 niños (72,4%) muestran una representación excesiva de los verbos (es decir, sobre la diagonal; media = 0,03). Estos 21 niños se superponen perfectamente, mostrando tanto una representación insuficiente de los sustantivos como una representación excesiva de los verbos. Entre los 8 niños restantes, 7 muestran el patrón inverso (es decir, NP/RR hipotético), aunque con un efecto mucho menor: sobrerrepresentación de sustantivos (media = 0,004) y subrepresentación de verbos (media = -0,008). A un niño no se le observaron raíces de sustantivos o verbos en su registro, por lo que obtuvo un puntaje de 0 en todas las medidas (Figura 3). Estas tendencias numéricas, si bien son consistentes, tienen por sí solas un valor inferencial débil. Lo que necesitamos para complementarlos son estimaciones iniciales del área por encima y por debajo de la diagonal a

lo largo de nuestra muestra para comprobar si es probable que lo que hemos observado sea diferente del azar.

Nuevamente seguimos a Braginsky y colegas (2019; Frank et al., 2021) al generar intervalos de confianza de la diferencia del aprendizaje imparcial en el intervalo del 0 al 100% del tamaño del vocabulario (Figura 4) mediante un muestreo aleatorio de los datos 1000 veces con reemplazo y luego ajustando el mismo modelo para cada muestra aleatoria para medir el área de sesgo aparente. Estas estimaciones de área se utilizaron luego para generar intervalos de confianza superior e inferior de una diferencia sin sesgo (es decir, un área de 0). En términos sencillos, las distribuciones con intervalos de confianza que se superponen en gran medida a la estimación cero son más consistentes con la falta de sesgo; los valores inferiores a cero indican una representación insuficiente y los valores superiores a cero indican una representación excesiva.

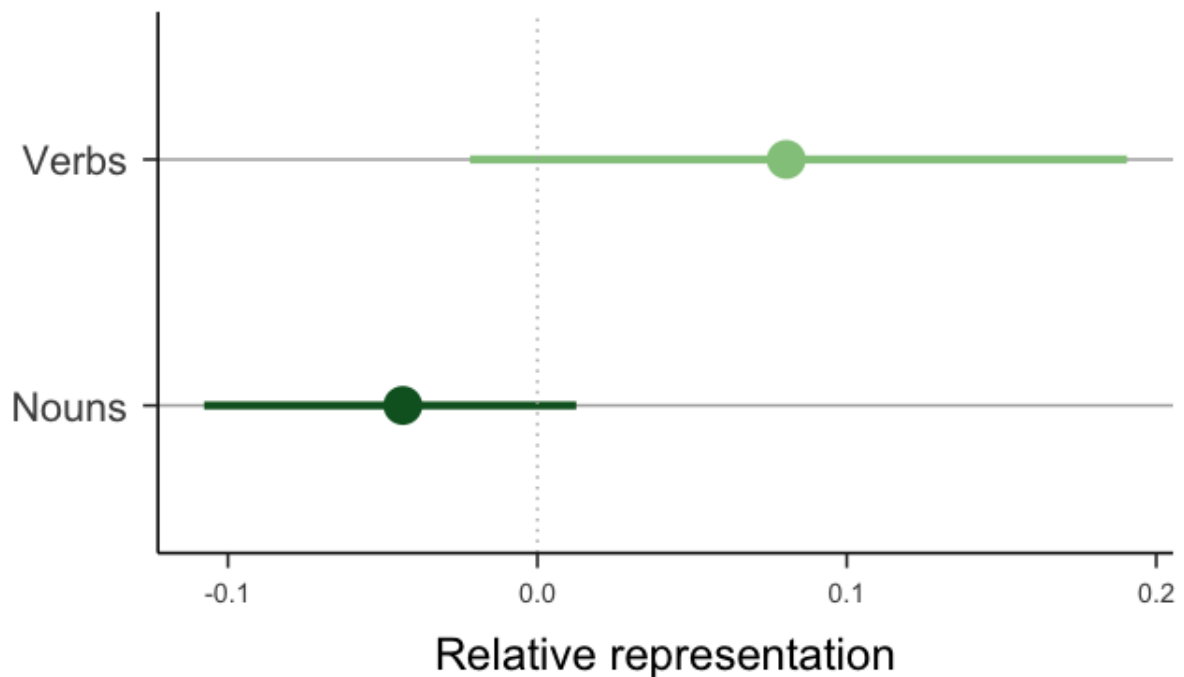


Figura 4. Estimaciones del efecto de representación relativa para sustantivos y verbos, con líneas horizontales gruesas que muestran los intervalos de confianza del 95 % con arranque para cada clase.

Si bien las estimaciones numéricas de los intervalos de confianza iniciados siguen siendo las mismas que las del análisis del conjunto de datos completo anterior (es decir, negativas para los sustantivos y positivas para los verbos), la superposición sustancial con cero en ambas distribuciones sugiere que no hay evidencia significativa de aprendizaje sesgado de sustantivos o verbos (área media para sustantivos: -0,043, IC del 95% [-0,108, 0,013]; verbos: 0.080, IC del 95% [-0,022, 0,191]; Figura 4). Observamos que la representación relativa estima los tamaños del efecto a partir de estos datos (sustantivos: -0,043; verbos: 0.080) son comparables a los casos estadísticamente significativos de sobrerepresentación y subrepresentación encontrados en Braginsky, Frank y colegas (2019; Frank et al., 2021); sin embargo, en nuestro conjunto de datos más pequeño y ruidoso, la incertidumbre estimada es mayor., especialmente dado que no tenemos datos observados para el intervalo de tamaño de vocabulario entre 0,21 y 1,0.

4. Discusión

El desarrollo temprano del lenguaje está impulsado en parte por los sesgos cognitivos que los niños aportan al proceso de aprendizaje. Una gran cantidad de evidencia (observacional, experimental y basada en listas de verificación) sugiere que los objetos concretos tienen una ventaja perceptiva y conceptual en el aprendizaje temprano de palabras en comparación con las relaciones, un patrón de desarrollo de vocabulario conocido como “sesgo del sustantivo”. En este artículo hemos examinado la hipótesis del sesgo sustantivo adquiriendo tseltal niños. Estábamos interesados en estudiar el desarrollo del vocabulario tseltal en particular porque las propiedades gramaticales del idioma tseltal y las prácticas comunicativas en la interacción niño-cuidador tseltal aumentan la prominencia conceptual de los verbos al tiempo que reducen la prominencia conceptual de los sustantivos. Trabajos anteriores que investigaban un sesgo sustantivo en tseltal habían parecido mixtos, con un resultado a favor del sesgo (Brown et al., 2005; Gentner, 2006; datos de lista de verificación) y otro en contra (Brown, 1998; habla espontánea), ambos usando pequeñas muestras, pero este último utiliza observación longitudinal en profundidad (ver también trabajo estrechamente relacionado de De León, 1999a,b; 2001). Ambos estudios habían sugerido un sesgo sustantivo translingüística mente débil y un apoyo más sólido a la noción de aprendizaje verbal temprano y rápido en tseltal. Los presentes hallazgos se basan en una muestra transversal más grande de habla tseltal altamente naturalista, que analizamos utilizando dos enfoques: un enfoque tradicional basado en medidas de frecuencia proporcional y un enfoque más novedoso que examina la representación relativa dentro de las clases de palabras.

Utilizando el método más tradicional (proporción de sustantivos y proporción de sustantivo-verbo), encontramos de hecho un sesgo numérico de sustantivo en ambas medidas, pero el sesgo fue pequeño (es decir, cerca de 0,5 en proporción: 0,56; cerca de 1,0 en proporción sustantivo-verbo: 1,48). Comparando lo mejor posible con trabajos anteriores, nuestras estimaciones actuales son similares pero ligeramente más favorables al sesgo sustantivo que los trabajos anteriores sobre lenguas mayas. Nuestras estimaciones actuales también están menos sesgadas que los seis idiomas informados en el artículo fundacional de Gentner (1982). Si bien es imposible hacer una comparación de manzanas con manzanas entre estos estudios, dada su variación en el tamaño de la muestra, el enfoque, el contexto de registro, el preprocesamiento de datos y más, los valores ampliamente similares logrados en las tres muestras de estudios tseltales existentes (proporciones de tipos de sustantivos de 0,53 a 0,56; proporciones de sustantivos a verbos de 1,17 a 1,48) aumentan nuestra confianza en la presencia de una dominancia proporcional pequeña pero consistente de los sustantivos encima verbos en el desarrollo temprano del vocabulario tseltal. Este es un patrón que no se puede atribuir únicamente al método (es decir, habla espontánea versus lista de verificación). Dicho esto, las medidas de frecuencia proporcional no tienen en cuenta las cantidades básicas de sustantivos y verbos que se deben aprender y, por lo tanto, no pueden proporcionar información directa sobre el aprendizaje imparcial versus el sesgado. Para ello, necesitamos una medida de representación relativa de las clases de palabras en comparación con el tamaño total del vocabulario, como lo hicieron anteriormente Bates, Braginsky, Frank y sus colegas (Bates et al., 1994; 1995; Braginsky et al., 2019; Frank et al., 2021).

Cuando utilizamos este método más novedoso (la representación relativa), no encontramos evidencia numérica o estadística de una sobrerrepresentación de los sustantivos; más bien vemos una tendencia numérica hacia la sobrerrepresentación de los verbos, un patrón que se mantiene en casi tres cuartas partes de las grabaciones individuales muestreadas. En otras palabras, Cuando nuestra medida del aprendizaje sesgado tiene en cuenta el tamaño de las categorías de sustantivos y verbos en el léxico que se va a adquirir, no vemos evidencia de una sobrerrepresentación de los sustantivos, lo que va en contra de la idea

del sesgo del sustantivo. Recientemente se utilizó un enfoque analítico lógicamente similar de representación relativa para mostrar un sesgo sustantivo constante en los vocabularios productivos informados por los cuidadores de niños que aprenden otros 27 idiomas (Frank et al., 2021), dando crédito a la idea de que nuestros resultados actuales no son simplemente por el enfoque analítico. Es importante destacar que somos los primeros (hasta donde sabemos) en intentar una adaptación del enfoque de representación relativa a los datos de transcripción, y una diferencia práctica importante al realizar el análisis es que sólo una pequeña fracción del vocabulario productivo de los niños se observará en las grabaciones. Según nuestros datos, el niño con el inventario más grande utilizó el 21% del total de raíces de sustantivos y verbos observados, en lugar del casi 100% de las palabras que podríamos esperar de una lista de verificación de vocabulario normado. Nuestro modelo restringido se ajusta a todo el espacio de 0 a 100 %, por lo que, de manera conservadora, va en contra de nuestra capacidad de evidenciar estadísticamente un sesgo, ya que no se observa ninguna diferencia con respecto al aprendizaje imparcial en el 80 % del ajuste total del inventario. Dicho esto, los datos de los registros individuales dan una idea clara de lo que podríamos ver con una mejor estimación del vocabulario acumulativo de los niños: una subrepresentación consistente y mayor de los sustantivos y una sobrerrepresentación de los verbos. Lo ideal sería volver a examinar estos hallazgos en trabajos futuros en los que (1) se observa una mayor cantidad de datos por niño para examinar mejor el espectro completo del desarrollo del vocabulario antes de los 52 meses y (2) se utilice una lista de verificación de vocabulario validada para probar el enfoque dentro de su contexto analítico original previsto.

Nuestros hallazgos combinados sugieren que la evidencia de un sesgo sustantivo en tseltal es muy débil: solo aparece cuando no se tienen en cuenta los tamaños de referencia del inventario y, aun así, los resultados del sesgo se encuentran entre los más bajos reportados en la literatura translingüística. Desde la perspectiva de la representación relativa, el patrón clásico de sesgo sustantivo se invierte numéricamente y aparece en cambio como un sesgo verbal, aunque críticamente los datos son demasiado escasos para producir inferencias estadísticas sólidas sobre el patrón de todo el grupo. Esta falta general de evidencia sólida de un sesgo sustantivo es esperable para el tseltal, dado que el Los patrones gramaticales e interaccionales del input lingüístico tseltal conducen a una mayor accesibilidad lingüística y conceptual de los verbos y una accesibilidad reducida de los sustantivos (Brown, 1998; de León, 1999a,b;2001; Polian, 2020). En términos más generales, el hallazgo de un sesgo sustantivo muy débil está en línea con otros trabajos comparativos interlingüísticos que muestran una amplia variación entre idiomas en el tamaño aparente del sesgo (p. ej., Au et al., 1994; Braginsky et al., 2019; Choi y Gopnik, 1995; Choi, 2000; Gopnik et al., 1996; Gopnik y Choi, 1990; Frank et al., 2021; Gentner, 1982; 2006; Gentner y Boroditsky 2001, 2008; Imai et al., 2008; Kim et al., 2000; Setoh et al., 2021; Tardif, 1996; Tardif et al., 1997; 1999; 2008; Xuan & Dollaghan, 2013).

Entonces, ¿qué nos dicen estos hallazgos sobre los sesgos cognitivos que impulsan el aprendizaje de palabras? La hipótesis NP/RR se centra en sustantivos objeto (especialmente animar unos) como óptimo ejemplares de accesibilidad conceptual (Gentner 1982;2006). Nuestros hallazgos, que siguen los de Brown, de León y otros (Brown, 1998; de León 1999a,b; 2001) sugieren una adición importante a esta historia: La accesibilidad conceptual en el momento depende no sólo de las características intrínsecas del referente, sino también de otros factores críticos como la estructura lingüística, la socialización, la atención culturalizada, la señalización multimodal, lo reciente, etc. Estos factores pueden desempeñar un papel importante al iluminar los referentes previstos y sus etiquetas lingüísticas previstas, dando (quizás momentáneamente) mapas conceptuales-lingüísticos bien definidos de conceptos confusos que de otro modo serían relativamente inaccesibles. Este mismo punto de vista puede ayudar a explicar el hecho de que el sesgo sustantivo es aparentemente más débil en algunas comunidades lingüísticas y culturales que en otras (Choi y Gopnik, 2005; Rosemberg et al. 2020; Snedeker et al., 2003): innumerables factores pueden directa o indirectamente entrenar

indirectamente la atención de los niños a categorías conceptuales relevantes, en el momento y en todas las instancias, preparándonos para el etiquetado lingüístico.

En este sentido, podríamos esperar ver efectos de accesibilidad que abarquen clases sintácticas, de modo que, si bien los sustantivos pueden tener una ventaja general de aprendizaje, se adquieren conceptos más accesibles desde el punto de vista socio interaccional, lingüístico y cultural antes que otros en una variedad de palabras diferentes. dominios (por ejemplo, sustantivos concretos y animados frente a abstractos; conceptos perceptualmente destacados frente a no salientes codificados en verbos). De hecho, Gentner (1982; 2006) predice específicamente este efecto dentro de la categoría de sustantivos y otros lo predicen dentro de la categoría de verbos (Brown, 1998, 2008; E. V. Clark, 1993; Ninio, 1996). Por lo tanto, aquí simplemente proponemos ampliar el alcance y la generalidad de las ventajas de la accesibilidad conceptual a cualquier factor que pueda aumentar significativamente la accesibilidad momentánea durante el aprendizaje de palabras (es decir, no solo la limitación conceptual inherente). De hecho, algo parecido a estos efectos de accesibilidad generalizados ya se captura en el trabajo de Braginsky, Frank y sus colegas (Braginsky et al., 2019; Frank et al., 2021), quienes encuentran que la concreción y la “infantilidad” de una palabra (es decir, fuerza de la relevancia del bebé) son predictores muy sólidos de qué tan temprano se informa que un niño comienza a producir esa palabra.

Por lo tanto, si tuviéramos más datos tseltales (habla espontánea o lista de verificación), todavía esperaríamos ver que, por ejemplo, dentro de cada categoría gramatical, habría efectos de concreción y, entre los sustantivos, una adquisición más temprana de palabras objeto concretas y animadas. Es decir, estos datos aún pueden ser parcialmente consistentes con la Hipótesis NP/RR, pero hay más que sugieren que los sesgos específicos de un objeto deben considerarse en equilibrio con los factores de accesibilidad de la educación lingüística y cultural del niño que influyen colectivamente en la trayectoria del desarrollo léxico temprano.

4.1 Limitaciones y direcciones futuras

Instamos a tener precaución al generalizar estos hallazgos por varias razones importantes. En primer lugar, si bien este conjunto de datos representa una muestra sin precedentes del habla de los niños tseltales (p. ej., casi triplicando el tamaño de la muestra en Casillas et al., 2020), el número de participantes en relación con el rango de edad aún es pequeño y los datos son transversales (no longitudinales como los otros trabajos sobre tseltal y Sobre Tsotsi; Marrón, 1998; 2009; de León, 1999a, 1999b, 2001). También hay solo 45 minutos de grabación observados para cada niño, y los clips fueron seleccionados al azar., que se manifestó como una gama muy limitada de vocabularios productivos observados en los niños. Sería ideal, por ejemplo, acercarnos más al habla de los niños que se encuentran en las primeras etapas de producción léxica entre nuestra amplia muestra de edad. Ya estamos planeando realizar un muestreo más denso de estas grabaciones; cuando estén listos, estos datos adicionales aumentarán la confiabilidad de nuestras estimaciones de léxico para cada niño y, en el proceso, reducirán la incertidumbre de nuestras estimaciones iniciales.

En segundo lugar, el análisis de estimación de representación relativa se diseñó originalmente para datos de listas de verificación normadas, Y hasta que podamos comparar nuestras estimaciones basadas en transcripciones con los datos de la lista de verificación, nuestros análisis actuales seguirán siendo algo experimentales.. Brown y colegas (2005) inicializar una lista de verificación de vocabulario para tseltal, pero requiere mayor refinamiento, pilotaje y análisis antes de que podamos usarla con una muestra más grande de niños tseltales para probar la representación relativa con datos de la lista de verificación (tal esfuerzo está en marcha; Autores, en progreso). A corto plazo, se podría abordar esta limitación replicando los análisis actuales con datos de sustantivos y verbos de grabaciones de un día en inglés para ver

si replicamos el hallazgo clásico de sesgo de sustantivos para el inglés (un idioma muy sesgado; Frank et al., 2021). o si hay algo en la representación relativa de los datos de transcripción que socava un aparente sesgo nominal. Finalmente, para completar completamente estos datos, idealmente también examinaremos experimentalmente el aprendizaje de palabras novedosas (p. ej., Childers & Tomasello, 2002; Imai et al., 2008), pero de una manera que sea sensible a los contextos culturalmente típicos para el aprendizaje interactivo de palabras en Tzeltal. Por ahora, tomamos estos hallazgos como un dato importante en el debate sobre el sesgo de los sustantivos y como apoyo a la perspectiva etnográficamente fundamentada sobre la adquisición de la lengua maya, que consistentemente ha enfatizado la fuerza del aprendizaje temprano de los verbos sobre el aprendizaje temprano de los sustantivos (de León, 1997; Brown, 1997; 1998; 2001; 2007; 2008; Brown et al., 2005; Pye et al., 2007).

4.2 Conclusión

Este estudio examinó la evidencia de un sesgo sustantivo en los vocabularios productivos de 29 adquiriendo tzeltal niños (de 9 a 52 meses) utilizando dos enfoques analíticos: frecuencia proporcional y representación relativa. Los niños que aprenden tzeltal encuentran una variedad de características lingüísticas y socio interactivas que disminuyen la prominencia de los sustantivos y al mismo tiempo aumentan la prominencia de los verbos, lo que motiva el presente estudio. Las estimaciones de frecuencia proporcional sugieren un sesgo sustantivo consistente, pero muy pequeño, que está en línea con trabajos anteriores informados sobre el desarrollo del vocabulario maya y menor que lo que se ha informado para otros idiomas. Las estimaciones de representación relativa, por otro lado, no sugieren ningún sesgo de sustantivos, y la mayoría de los niños muestran datos invertidos numéricamente: sobrerrepresentación de verbos y subrepresentación de sustantivos. Nuestros hallazgos hacen eco de investigaciones transculturales pasadas en este dominio: la influencia de la accesibilidad conceptual inherente de un referente (a la hipótesis NP/RR) está mediada sustancialmente por factores lingüísticos y culturales que pueden causar un debilitamiento (a veces una aparente reversión) de su predicción. Efectos sobre el aprendizaje temprano de palabras.

Referencias citadas

- Akhtar, N. (2005). La solidez del aprendizaje a través de la escucha. *Ciencia del desarrollo*, 8(2), 199–209. <https://doi.org/10.1111/j.1467-7687.2005.00406.x>
- Arunachalam, S. (2013). Los niños de dos años pueden empezar a adquirir significados verbales en contextos socialmente empobrecidos. *Cognición*, 129(3), 569–573. <https://doi.org/10.1016/j.cognition.2013.08.021>
- Au, TKF, Dapretto, M. y Song, YK (1994). Entradas versus restricciones: adquisición temprana de palabras en coreano e inglés. *Revista de Memoria y Lenguaje*, 33(5), 567–582.
- Aust, F., Barth, M. (2022). papaya: Prepare artículos de revistas APA reproducibles con R Markdown. Versión del paquete R 0.1.1, <https://github.com/crsh/papaja>
- Baillargeon, R. (1993). El concepto de objeto revisitado: nuevas direcciones en la investigación del conocimiento físico de los bebés. En C.E. Ganrud (Ed.) *Simposios Carnegie Mellon sobre cognición, vol. 23: Percepción visual y cognición en la infancia*. (págs. 265–315). Hillsdale, Nueva Jersey: Erlbaum. <https://doi.org/10.4324/9780203763032>
- Bassano, D. (2000). Desarrollo temprano de sustantivos y verbos en francés: exploración de la interfaz entre léxico y gramática. *Revista de Lenguaje Infantil*, 27(3), 521–559. doi:10.1017/S0305000900004396
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, JS, et al. (1994). Variación evolutiva y estilística en la composición del vocabulario temprano. *Revista de lenguaje infantil*, 21, 85-123. doi:10.1017/S0305000900008680
- Bates, E., Dale, PS y Thal, D. (1995). Diferencias individuales y sus implicaciones para las teorías del desarrollo del lenguaje. En P. Fletcher y B. MacWhinney (Eds.), *El manual del lenguaje infantil*, 30 (págs. 95-151). Oxford, Reino Unido: Blackwell. <https://doi.org/10.1111/b.9780631203124.1996.00005.x>
- Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Ajuste de modelos lineales de efectos mixtos utilizando lme4. *Revista de software estadístico*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L.,... & Vyt, A. (2004). Análisis translingüístico del vocabulario en niños pequeños: español, holandés, francés, hebreo, italiano, coreano e inglés americano. *Desarrollo infantil*, 75(4), 1115-1139.
- Braginsky, M., Yurovsky, D., Marchman, V. A. y Frank, M. C. (2019). Coherencia y variabilidad en el aprendizaje de palabras de los niños en todos los idiomas. *Mente abierta*, 3, 52–67. https://doi.org/10.1162/opmi_a_00026
- Marrón, P. (1997). Aislado la raíz CVC en maya tzeltal: un estudio de los primeros verbos infantiles. En EV Clark (Ed.), *Actas del 28.º Foro Anual de Investigación del Lenguaje Infantil* (págs. 41 a 52). Stanford, CA: CSLI/Prensa de la Universidad de Chicago.
- Marrón, P. (1998). Verbos tzeltales tempranos: estructura argumental y representación argumental. En E. Clark (Ed.), *Actas del 29º Foro Anual de Investigación del Lenguaje Infantil de Stanford* (págs. 129-140). Stanford: Publicaciones CSLI. <http://hdl.handle.net/11858/00-001M-0000-0013-3DB7-8>
- Marrón, P. (2001). Aprender a hablar sobre el movimiento ARRIBA y ABAJO en tzeltal: ¿Existe un sesgo específico del idioma para el aprendizaje de verbos? En M. Bowerman y S. C. Levinson (Eds.), *Adquisición del lenguaje y desarrollo conceptual*. (págs. 512 a 543). Cambridge: Prensa de la Universidad de Cambridge. <https://doi.org/10.1017/CBO9780511620669>
- Marrón, P. (2007). Influencias específicas de la cultura en el desarrollo semántico Adquirir la construcción tzeltal 'benefactiva'. En B. Pfeiler (Ed.), *Aprendizaje de lenguas indígenas: Adquisición de lenguas infantiles en Mesoamérica* (págs. 119-154). Mouton de Gruyter: Berlín. <https://doi.org/10.1515/9783110923148>

- Marrón, P. (2008). Especificidad verbal y realización de argumentos en lengua infantil tzeltal. En M. Bowerman y P. Brown (Eds.), *Perspectivas translingüísticas sobre la estructura argumental: implicaciones para la capacidad de aprendizaje* (págs. 167–189). Mahwah, Nueva Jersey: Erlbaum.
- Marrón, P. (2011). La organización cultural de la atención. En A. Duranti, E. Ochs y B. B. Schieffelin (Eds.), *El manual de socialización lingüística* (págs. 29 a 55). Malden, MA: Wiley-Blackwell. [doi:10.1002/9781444342901](https://doi.org/10.1002/9781444342901)
- Marrón, P. (2014). El contexto interaccional del aprendizaje de lenguas en tzeltal. En I. Arnón, M. Casillas, C. Kurumada y B. Estigarriba (Eds.), *Lenguaje en interacción: estudios en honor a Eve V. Clark* (págs. 51 a 82). Ámsterdam: Benjamins. <https://doi.org/10.1075/tilar.12>
- Brown, P., Gentner, D. y Braun, K. (julio de 2005) Adquisición inicial de sustantivos y verbos en niños que hablan tzeltal. *Décimo Congreso Internacional para el Estudio del Lenguaje Infantil*. Berlín, Alemania.
- Brown, P., Pfeiler, B., de Leon, L. y Pye, C. (2013). La adquisición del acuerdo en cuatro lenguas mayas. En E. Bavin y S. Stoll (Eds.), *La adquisición de la ergatividad* (págs. 271–306). Ámsterdam: Benjamins.
- Bunce, J. P., Soderstrom, M., Bergelson, E., Rosemberg, C. R., Stein, A., Alam, F.,... Casillas, M. (2 de septiembre de 2020). Un examen transcultural de las experiencias lingüísticas cotidianas de los niños pequeños. <https://doi.org/10.31234/osf.io/723pr>
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L. y Weir, J. (1995). Un estudio translingüístico del desarrollo léxico temprano. *Desarrollo cognitivo*, 10(2), 159-199.
- Casillas, M. (2023). Proyecto: Investigación del sesgo sustantivo en la adquisición temprana del lenguaje.
- Casillas, M., Brown, P. y Levinson, SC (2017). Corpus de Casillas HomeBank. [doi:10.21415/T51X12](https://doi.org/10.21415/T51X12)
- Casillas, M., Bergelson, E., Warlaumont, A. S., Cristia, A., Soderstrom, M., VanDam, M. y Sloetjes, H. (2017). Un nuevo flujo de trabajo para anotaciones semiautomatizadas: pruebas con grabaciones naturalistas de formato largo de entornos lingüísticos de niños. En *Actas de Interspeech 2017* (págs. 2098-2102). [doi:10.21437/Interspeech.2017-1418](https://doi.org/10.21437/Interspeech.2017-1418).
- Casillas, M., Brown, P. y Levinson, SC (2020). Experiencia lingüística temprana en un pueblo maya tseltal. *Desarrollo infantil*, 91(5), 1819–1835. <https://doi.org/10.1111/cdev.13349>
- Chang-Song, YK (1997). Adquisición temprana de palabras en coreano: evidencia de restricciones Ver (I). *Revista Coreana de Desarrollo Humano*, 4(1), 76–86.
- Childers, JB y Tomasello, M. (2002). Los niños de dos años aprenden nuevos sustantivos, verbos y acciones convencionales mediante exposiciones masivas o distribuidas. *Psicología del desarrollo*, 38(6), 967–978. <https://doi.org/10.1037/0012-1649.38.6.967>
- Childers, JB, Porter, B., Dolan, M., Whitehead, C. B. y McIntyre, K. P. (2020). ¿La atención visual de los niños a objetos específicos afecta su aprendizaje verbal? *Primer idioma*, 40(1), 21–40. <https://doi.org/10.1177/0142723719875575>
- Childers, JB, Warkentin, E., Porter, B. M., Young, M., Lalani, S. y Gopalkrishnan, A. (2022). El procesamiento de eventos por parte de niños en edad preescolar durante el aprendizaje de verbos: ¿el foco está en las personas (rostros) o en sus acciones (manos)? *Ciencias del cerebro*, 12(3), 344. <https://doi.org/10.3390/brainsci12030344>
- Choi, S. (2000). Aportes del cuidador en inglés y coreano: uso de sustantivos y verbos en contextos de lectura de libros y juegos con juguetes. *Revista de lenguaje infantil*, 27(1), 69–96. [doi:10.1017/S0305000999004018](https://doi.org/10.1017/S0305000999004018)
- Clark, EV (1995). El léxico en la adquisición (núm. 65). Prensa de la Universidad de Cambridge.

- Clark, HH (1996). Usando el lenguaje. Prensa de la Universidad de Cambridge.
- De León, L. (1999a). Raíces verbales y habla del cuidador en la adquisición temprana del maya tzotzil. En B. Box, D. Juravsky y L. Michaelis (Eds.), *Cognición, discurso y función*. (págs. 99-119). Stanford, CA: CSLI/Prensa de la Universidad de Chicago.
- De León, L. (1999b.) Verbos en tzotzil desarrollo sintáctico temprano. *Revista Internacional de Bilingüismo*, 3(2/3): 219–240. <https://doi.org/10.1177/13670069990030020601>
- De León, L. (2001). Encontrar el camino más rico: El papel del lenguaje y la cognición en la adquisición de la verticalidad en tzotzil. En M. Bowerman y S. Levison (Eds.), *Adquisición del lenguaje y desarrollo conceptual*. (págs. 544–565). Cambridge: Prensa de la Universidad de Cambridge.
- Fenson, L. (2007). Inventarios de desarrollo comunicativo MacArthur-Bates. Baltimore, MD: Compañía editorial Paul H. Brookes.
- Frank, M. C., Braginsky, M., Yurovsky, D. y Marchman, V. A. (2021). *Variabilidad y coherencia en el aprendizaje temprano de idiomas: el proyecto Wordbank*. Cambridge, MA: MIT Press.
- Gentner, D. (1982). Por qué los sustantivos se aprenden antes que los verbos: relatividad lingüística versus partición natural. *Informe Técnico No. 257, Centro de Estudios de la Lectura*. Washington D. C.: Instituto Nacional de Educación. <https://eric.ed.gov/?id=ED219724>
- Gentner, D. (2006). Por qué los verbos son difícil de aprender. En K. Hirsh-Pasek y R. Golinkoff (Eds.) *La acción se encuentra en la palabra: Cómo aprenden los niños los verbos* (págs. 544–564). Prensa de la Universidad de Oxford. [doi:10.1093/acprof:oso/9780195170009.001.0001](https://doi.org/10.1093/acprof:oso/9780195170009.001.0001)
- Gentner, D. y Boroditsky, L. (2001). Individuación, relatividad y aprendizaje temprano de palabras. En M. Bowerman y S. C. Levinson (Eds.), *Adquisición del lenguaje y desarrollo conceptual*. (págs. 215-256). Cambridge: Prensa de la Universidad de Cambridge. <https://doi.org/10.1017/CBO9780511620669>
- Gentner, D. y Boroditsky, L. (2008). Adquisición temprana de sustantivos y verbos: evidencia de Navajo. En V. C. Mueller Gathercole (Ed.), *Rutas hacia el lenguaje: Estudios en honor a Melissa Bowerman* (págs. 35 a 68). Londres, Reino Unido: Psychology Press.
- Gillette, J., Gleitman, H., Gleitman, L. y Lederer, A. (1999). Simulaciones humanas de aprendizaje de vocabulario. *Cognición*, 73(2), 135-176. [https://doi.org/10.1016/S0010-0277\(99\)00036-0](https://doi.org/10.1016/S0010-0277(99)00036-0)
- Gleitman, L. (1990). Las fuentes estructurales de los significados verbales. *Adquisición de lenguaje*, 1(1), 3–55. https://doi.org/10.1207/s15327817la0101_2
- Gopnik, A. y Choi, S. (1990). ¿Las diferencias lingüísticas conducen a diferencias cognitivas? Un estudio translingüístico del desarrollo semántico y cognitivo. *Primer idioma*, 10(30), 199-215. <https://doi.org/10.1177/014272379001003002>
- Gopnik, A., Choi, S. y Baumberger, T. (1996). Diferencias interlingüísticas en el desarrollo cognitivo y semántico temprano. *Cognitivo y desarrollo*, 11(2), 197–225.
- Hammarström, H. y Forkel, R., Haspelmath, M. y Bank, S. (2023). Glotología 4.8. Leipzig: Instituto Max Planck de Antropología Evolutiva. <https://doi.org/10.5281/zenodo.8131084> Consultado: 2023-08-23.
- Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M. y Shigematsu, J. (2008). Aprendizaje novedoso de sustantivos y verbos en niños de habla china, inglesa y japonesa. *Desarrollo infantil*, 79(4), 979–1000. <https://doi.org/10.1111/j.1467-8624.2008.01171.x>
- Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E. y Gutiérrez-Clellen, V. (1993). Desarrollo léxico temprano en bebés y niños pequeños de habla hispana. *Diario del idioma infantil*, 20(3), 523–549.

- Kim, M., McGregor, K. y Thompson, C. (2000). Desarrollo léxico temprano en niños de habla inglesa y coreana: patrones de lengua general y específicos de la lengua. *Revista de lenguaje infantil*, 27(2), 225–254. [doi:10.1017/S0305000900004104](https://doi.org/10.1017/S0305000900004104)
- Landau, B. y Gleitman, L. (1985). *Lenguaje y experiencia: evidencia del niño ciego*. Cambridge, MA: Harvard University Press.
- Levinson, SC (2019). Fundamentos interaccionales del lenguaje: la hipótesis del motor de interacción. En P. Hagoort (Ed.) *Lenguaje humano: de los genes y el cerebro al comportamiento* (págs. 189-200). Boston, MA: MIT Press.
- Marchman, V. A., Martinez-Sussmann, C. y Dale, P. S. (2004). La naturaleza específica del lenguaje del desarrollo gramatical: evidencia de estudiantes de idiomas bilingües. *Ciencia del desarrollo*, 7, 212–224. <https://doi.org/10.1111/j.1467-7687.2004.00340.x>
- Markman, EM (1989). *Categorización y denominación en niños: Problemas de inducción*. La prensa del MIT.
- Markman, EM (1990). Limitaciones que los niños imponen al significado de las palabras. *Cognición*, 14, 57–77. https://doi.org/10.1207/s15516709cog1401_4
- Matthew Peter, P. 2015. La adquisición de la inflexión en la pronunciación maya. Ámsterdam: John Benjamins.
- Moroz, G. (2021). phonfieldwork: herramientas de trabajo de campo fonético lingüístico. Versión del paquete R 0.0.11. <https://CRAN.R-project.org/package=phonfieldwork>
- Niño, A. (1999). Verbos pioneros en el desarrollo sintáctico y la cuestión de la transitividad prototípica. *Revista de lenguaje infantil*, 26(3), 619–653. [doi:10.1017/S0305000999003931](https://doi.org/10.1017/S0305000999003931)
- Página, S. (1993). *Vocabulario temprano en coreano: ¿son más fáciles de aprender los sustantivos que los verbos?*. Universidad de Kansas. Tesis doctoral inédita.
- Pfeiler, B., Bittner, D., Dressler, W. y Kilani-Schoch, M. (2003). Adquisición temprana del complejo verbal en maya yucateco. *Desarrollo de la inflexión verbal en la adquisición de la primera lengua*, 379–399.
- Polian, G. (2013). Gramática del Tseltal de Oxchuc. Mexico City: Centro de Investigaciones y Estudios Superiores en Antropología Social.
- Polian, G. (2020). Diccionario multidialectal tseltal-español. *Diccionario*, 10. 1–8109. [doi:10.5281/zenodo.5526550](https://doi.org/10.5281/zenodo.5526550)
- Pye, C. (1985). La Adquisición de la Transitividad en el maya quiché. Artículos e informes sobre el desarrollo del lenguaje infantil, 24, 115–22.
- Pye, C., Pfeiler, B., De León, L., Brown, P. y Mateo, P. (2007). ¿Raíces o bordes? Explicar la variación en las formas verbales tempranas de los niños en cinco idiomas mayas. En B. Pfeiler (Ed.), *Aprendizaje de lenguas indígenas: Adquisición de lenguas infantiles en Mesoamérica* (págs. 15 a 46). Berlín: Mouton de Gruyter. <https://doi.org/10.1515/9783110923148>
- Pye, C., Pfeiler, B. y Peter, PM (2017). Adquisición de la lengua maya. En *Las lenguas mayas* (págs. 19 a 42). Routledge.
- Rosemberg, C. R., Alam, F., Audisio, C. P., Ramirez, M. L., Garber, L. y Migdalek, M. J. (2020). Sustantivos y verbos en el entorno lingüístico de los niños pequeños argentinos: diferencias socioeconómicas y relacionadas con el contexto. *Primer idioma*, 40(2), 192–217. <https://doi.org/10.1177/0142723719901226>
- Schieffelin, BB (1979). Getting it together: una aproximación etnográfica al estudio del desarrollo de la competencia comunicativa. En E. Ochs y B. B. Schieffelin (Eds.), *Pragmática del desarrollo*. Nueva York: Academic Press.
- Schneider, RM, Yurovsky, D. y Frank, MC (2015). Investigaciones a gran escala sobre la variabilidad en las primeras palabras de los niños. En Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D. y Maglio, P. P. (Eds.) *Actas*

- de la 37.^a reunión anual de la Sociedad de Ciencias Cognitivas (págs. 2110-2115). Austin, TX: Sociedad de Ciencias Cognitivas.
- Setoh, P., Cheng, M., Bornstein, M. y Esposito, G. (2021). Sesgos léxicos contrastantes en el habla bilingüe inglés-mandarín: madres con sesgo verbal, pero niños pequeños con sesgo sustantivo. *Revista de lenguaje infantil*, 48 (6), 1185-1208. [doi:10.1017/S0305000920000720](https://doi.org/10.1017/S0305000920000720)
- Simons, G. F. y Fennig, C. D. (Eds.) (2017). *Ethnologue: Lenguas del mundo*, vigésima edición. Dallas, Texas: SIL Internacional. <http://www.ethnologue.com>. Consultado: 2023-08-23.
- Skirgard, H. (2017). Humanos que leen gramáticas: la mitad de los idiomas del mundo no son indoeuropeos (pero casi la mitad de la población habla un idioma indoeuropeo). <http://humans-who-read-grammars.blogspot.com/2017/03/half-of-worlds-languages-are-not-indo.html> Consultado: 2023-08-23.
- Snedeker, J., Geren, J. y Shafto, C. (2012). Desenredando los efectos del desarrollo cognitivo y la experiencia lingüística: un estudio longitudinal de la adquisición del inglés en niños adoptados internacionalmente. *Psicología Cognitiva*, 65, 39–76.
- Snedeker, J., Li, P. y Yuan, S. (2003). Diferencias interculturales en el aporte al aprendizaje temprano de palabras. En *Actas de la reunión anual de la Sociedad de Ciencias Cognitivas*, 25, 1094–1099. Boston, MA: Sociedad de Ciencias Cognitivas. <https://escholarship.org/uc/item/15545365>
- Soderstrom, M., Casillas, M., Bergelson, E., Rosenberg, C., Alam, F., Warlaumont, A. S. y Bunce, J. (2021) Desarrollo de un sistema de anotación transcultural y MetaCorpus para estudiar la vida real de los bebés -Experiencia en idiomas mundiales. *colaboración*, 7(1), 23445. <https://doi.org/10.1525/collabra.23445>
- Soja, NN, Carey, S. y Spelke, ES (1991). Las categorías ontológicas guían las inducciones del significado de las palabras por parte de los niños pequeños: términos de objeto y términos de sustancia. *Cognición*, 38(2), 179–211. [https://doi.org/10.1016/0010-0277\(91\)90051-5](https://doi.org/10.1016/0010-0277(91)90051-5)
- Spelke, EA (1985). Percepción de unidad, persistencia e identidad: pensamientos sobre la concepción de los objetos por parte de los bebés. En J. Mehler y R. Fox (Eds.) *Cognición neonatal: más allá de la floreciente confusión* (págs. 89-114). Hillsdale, Nueva Jersey: Erlbaum.
- Spelke, EA (1990). Principios de percepción de objetos. *Ciencia cognitiva*, 14, 29–56. [https://doi.org/10.1016/0364-0213\(90\)90025-R](https://doi.org/10.1016/0364-0213(90)90025-R)
- Tardif, T. (1996). Los sustantivos no siempre se aprenden antes que los verbos: evidencia de los primeros vocabularios de los hablantes de mandarín. *Psicología del desarrollo*, 32(3), 492–504. <https://doi.org/10.1037/0012-1649.32.3.492>
- Tardif, T., Shatz, M. y Naigles, L. (1997). Habla del cuidador y uso de sustantivos versus verbos por parte de los niños: una comparación del inglés, el italiano y el mandarín. *Revista de Lenguaje Infantil*, 24(3), 535–565. [doi:10.1017/S030500099700319X](https://doi.org/10.1017/S030500099700319X)
- Tardif, T., Gelman, SA y Xu, F. (1999). Poner el "sesgo del sustantivo" en contexto: una comparación del inglés y el mandarín. *Desarrollo infantil*, 70(3), 620–635. <https://doi.org/10.1111/1467-8624.00045>
- Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N. y Marchman, VA (2008). Las primeras 10 palabras del bebé. *Psicología del desarrollo*, 44(4), 929.
- Waxman, SR (1991). Convergencias entre organización semántica y conceptual en los años preescolares. En J. P. Byrnes y SA Gelman (Eds.) *Perspectivas sobre lenguaje y pensamiento: interrelaciones en el desarrollo* (págs. 107-145). Cambridge, Reino Unido: Cambridge University Press. <https://doi.org/10.1017/CBO9780511983689>
- Waxman, SR y Hall, DG (1993). El desarrollo de un vínculo entre los sustantivos contables y las categorías de objetos: evidencia de bebés de quince a veinte meses. *Desarrollo infantil*, 64, 1224-1241. <https://doi.org/10.1111/j.1467-8624.1993.tb04197.x>

- Wickham H. (2016). *ggplot2: Gráficos elegantes para análisis de datos*. Springer-Verlag Nueva York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, TL, Miller, E., Bache, SM, Müller, K., Ooms, J., Robinson, D., Seidel, DP, Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019). Bienvenido al tidyverso. *Revista de software de código abierto*, 4(43), 1686. [doi:10.21105/joss.01686](https://doi.org/10.21105/joss.01686)
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. y Sloetjes, H. (2006). ELAN: un marco profesional para la investigación multimodal. En la Quinta Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC 2006) (págs. 1556-1559). <http://hdl.handle.net/11858/00-001M-0000-0013-1E7E-4>
- Xuan, L. y Dollaghan, C. (2013). Sesgo de sustantivo específico del idioma: evidencia de niños bilingües. *Revista de lenguaje infantil*, 40 (5), 1057-1075. [doi:10.1017/S0305000912000529](https://doi.org/10.1017/S0305000912000529)

Materiales complementarios

SM1. Alternativa de vocabulario básico: combinación de raíces producidas por el niño objetivo y discurso dirigido al niño objetivo

Nuestro análisis de representación relativa (consulte la sección 1.3 del texto principal para obtener una explicación y la sección 3.2 del texto principal para obtener resultados) define el vocabulario básico que se adquirirá sobre la base de todas las raíces nominales y verbales observadas en el habla dirigida al niño objetivo. Sin embargo, debido a que este es el primer intento de adaptar un método derivado de una lista de verificación de vocabulario a las transcripciones, aún no sabemos cuál es el mejor conjunto de expresiones para crear una lista de raíces de vocabulario básico. Una alternativa razonable a lo que informamos en el texto principal es utilizar raíces nominales y verbales que observamos tanto en el habla de los niños objetivo como en el discurso dirigido a los niños objetivo. En esta sección de Materiales complementarios revisamos cómo se verían los resultados si hubiéramos elegido esta definición alternativa para el vocabulario básico.

En primer lugar, y como era de esperar, la proporción máxima de tallos producidos por un niño individual es menor que la que vemos en el texto principal (menor porque el número total de tallos es mayor): aquí el máximo es 0,15 del inventario total, 0,14 de el inventario de sustantivos y 0,21 del inventario de verbos (en comparación con 0,21, 0,19 y 0,30, respectivamente, con el método producido únicamente por el niño objetivo en el texto principal). En segundo lugar, las tendencias numéricas son las mismas: los sustantivos muestran una diferencia numéricamente negativa con respecto al aprendizaje imparcial, mientras que los verbos muestran una diferencia numéricamente positiva. En tercer lugar, el mismo enfoque de arranque reportado en el texto principal revela que ambas clases de palabras se superponen sustancialmente con cero en sus distribuciones de área estimadas (área media para sustantivos: -0,034, *IC del 95%* [-0,129, 0,051]; verbos: 0,060, *IC del 95%* [-0,088, 0,215]). Al igual que en los resultados del texto principal, las áreas estimadas son comparables a las reportadas en los datos de la lista de verificación de Braginsky y colegas (2019; Frank et al., 2021), pero la variación entre las muestras bootstrap es mayor.

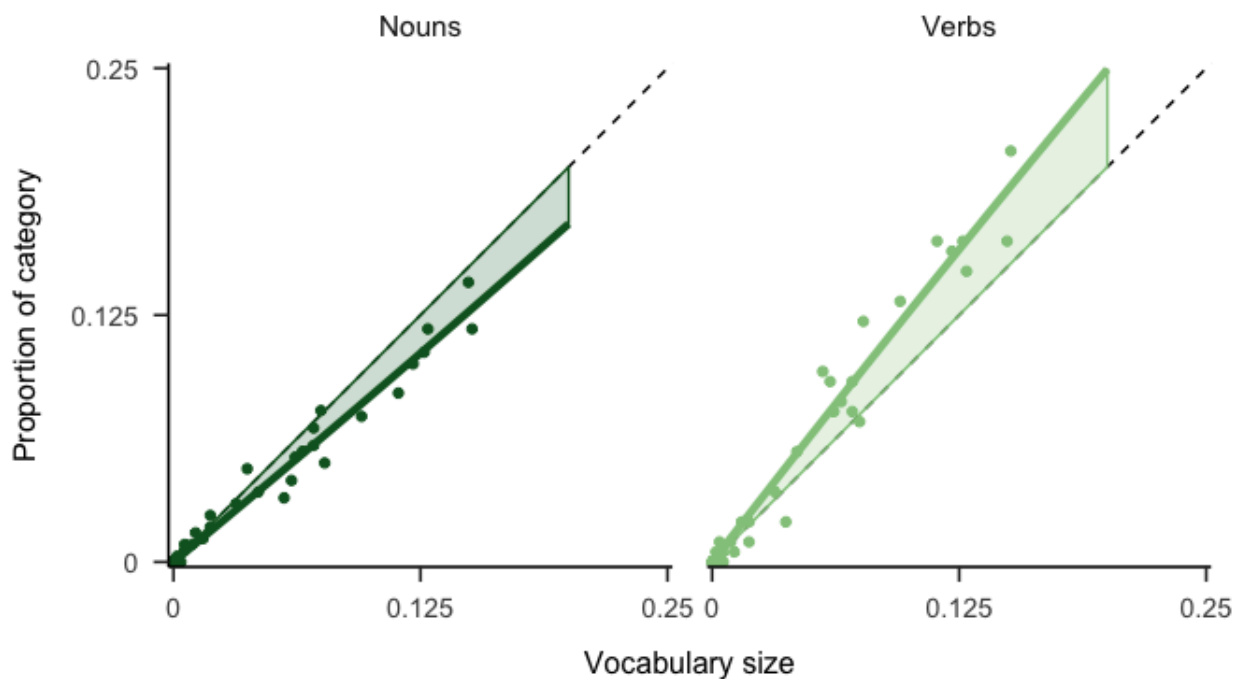


Figura SM1a. Representación relativa de sustantivos (izquierda) y verbos (derecha): proporción de raíces adquiridas para la categoría dada comparada con la proporción de raíces adquiridas para el inventario total de raíces de sustantivos y verbos. Cada punto de datos representa un niño individual y la línea gruesa muestra el ajuste restringido del modelo, con las regiones sombreadas resaltando el área estimada entre la trayectoria de desarrollo imparcial esperada en el caso de que no haya sesgo (la línea diagonal) y los datos reales observados (los puntos de datos).) en el rango de tamaños de vocabulario observados (tamaño relativo total del vocabulario = $[0,00:0,15]$). El vocabulario básico utilizado para este análisis de materiales complementarios incluye tanto el discurso producido por el niño objetivo como el dirigido por el niño objetivo, mientras que los análisis del texto principal solo incluyen el primero, con resultados cualitativamente equivalentes.

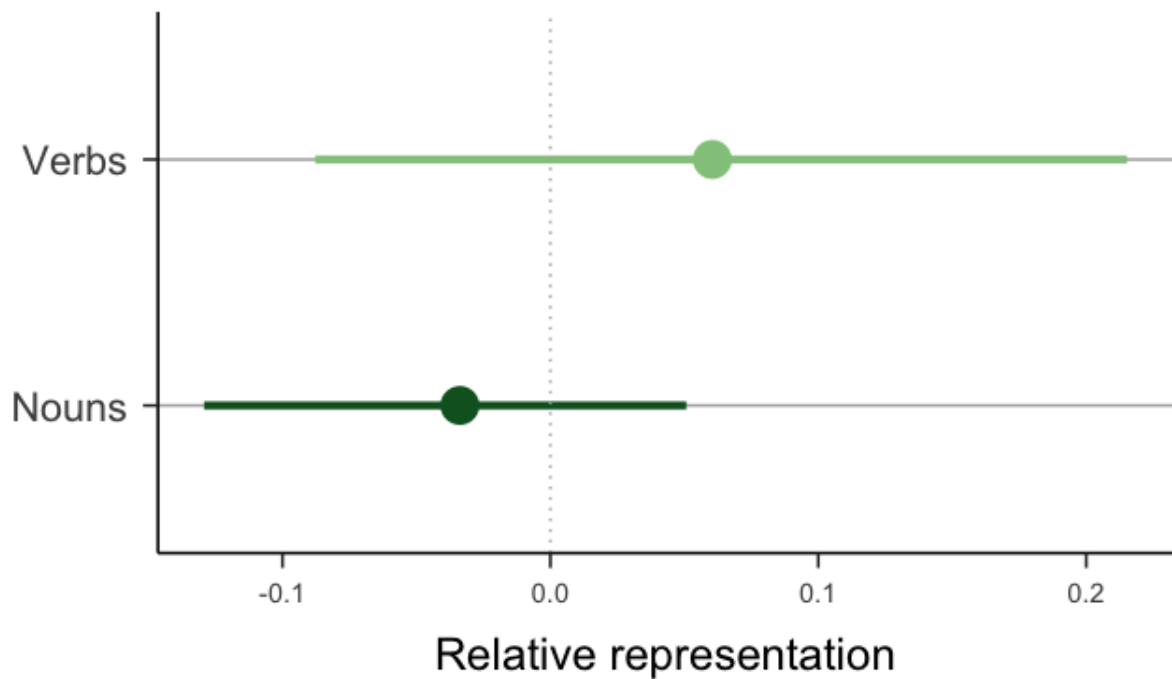


Figura SM1b. Estimaciones del efecto de representación relativa para sustantivos y verbos, con líneas gruesas horizontales que muestran los intervalos de confianza del 95 % con arranque para cada clase. El vocabulario básico utilizado para este análisis de materiales complementarios incluye tanto el discurso producido por el niño objetivo como el dirigido por el niño objetivo, mientras que los análisis del texto principal solo incluyen el primero, con resultados cualitativamente equivalentes.