

**DOSSIER**

**Género gramatical y género indexical en contacto: El uso de demostrativos en el español amazónico peruano**

Grammatical gender and indexical gender: demonstrative use in Peruvian Amazonian Spanish



*Rosa Vallejos-Yopán*

University of New Mexico, EE. UU.

rvallejos@unm.edu

ORCID: 0000-0001-7282-7611

*Marin Alemán-Ortiz*

University of New Mexico, EE. UU.

maleman10@unm.edu

ORCID: 0009-0004-4442-4195

Recibido: 07.07.25 — Aceptado: 24.10.25

<https://doi.org/10.30920/letras.96.144.5>



## RESUMEN

Este estudio examina el español hablado en la Amazonía de Perú para investigar el rol del contacto lingüístico en los patrones de variación en la concordancia de género entre los demostrativos y sus anáforas. El kukama y el español tienen sistemas de género tipológicamente diferentes que se extienden a los demostrativos. Mientras que el kukama posee género indexical que codifica el género de los participantes en el acto de habla, el español tiene género gramatical que categoriza a los sustantivos en clases definidas. En otras palabras, el kukama distingue categóricamente el género del hablante, y el español el género del referente. Para evaluar efectos de contacto, comparamos muestras de hablantes de español de comunidades kukama y de la ciudad de Iquitos. Considerando su perfil sociolingüístico, los participantes fueron organizados en tres grupos: monolingües en español, bilingües con kukama como L1 y español como L2, y hablantes de español como L1 y kukama como lengua de herencia. Entrevistas libres a 14 participantes produjeron un total de 1416 instancias de demostrativos que fueron codificadas en función de un conjunto de variables lingüísticas y extralingüísticas. Un análisis cuantitativo muestra que la concordancia de género es sensible a tres variables lingüísticas: el género del demostrativo, el rol sintáctico del demostrativo y la animacidad de la anáfora. También son significativas dos variables extralingüísticas: el perfil sociolingüístico y el género social del participante. Para explicar estos resultados, consideramos la interacción de los diferentes factores y los efectos del sustrato kukama en la reorganización del sistema de demostrativos en el español amazónico peruano. Estos hallazgos contribuyen al debate sobre el rol de la distancia tipológica entre las lenguas en el cambio lingüístico motivado por el contacto.

**PALABRAS CLAVE:** demostrativos; español amazónico; kukama; contacto lingüístico; lengua de herencia.

## ABSTRACT

This study examines Spanish spoken in the Peruvian Amazon to investigate contact effects in the patterns of variation in gender agreement between demonstratives and their anaphors. To evaluate contact effects, we compared samples of Spanish speakers from Kukama villages and the city of Iquitos. While Kukama has indexical gender that encodes the gender of the participants in the speech act, Spanish has grammatical gender that categorizes nouns into defined classes. Considering their sociolinguistic profile, participants were organized into three groups: a) Spanish monolinguals; b) Kukama/ Spanish bilinguals; and, c) Spanish L1 and Kukama as heritage language. Sociolinguistic interviews with 14 participants produced a total of 1 416 instances of demonstratives, which were coded for linguistic and extralinguistic variables. A quantitative analysis shows that gender agreement is sensitive to the demonstrative itself, its syntactic role, the animacy of the anaphora, as well as sociolinguistic profile and the social gender of the participants. To explain these results, we consider the interaction of different factors and the effects of the Kukama substratum on the reorganization of the demonstrative system in Peruvian Amazonian Spanish. These findings contribute to the debate on the role of typological distance between languages in language change driven by contact.

**KEYWORDS:** demonstratives; Amazonian Spanish; Kukama; language contact; heritage language.

## 1.1 Introducción

La deixis, el sistema con el que las lenguas vinculan sus enunciados con factores pragmáticos como el tiempo, el lugar y los participantes, muestra una rica variación interlingüística. Este estudio aborda el uso de demostrativos en el español amazónico peruano (EAP), una variedad de contacto (Vallejos, 2014), para examinar los resultados estructurales del encuentro de dos sistemas de género tipológicos diferentes. Más específicamente, analiza el habla de bilingües español-kukama para evaluar los posibles efectos en la codificación del género, dado que las lenguas involucradas tienen sistemas de género diferentes. El kukama (tupí) presenta indexicalidad de género (Vallejos, 2015), mientras que el español presenta género gramatical. Mientras que la indexicalidad de género codifica el género social de los participantes en el acto de habla (Levinson, 1983), el género gramatical codifica las clases nominales y se expresa mediante sistemas de concordancia (Corbett, 2007). En otras palabras, el kukama distingue el género de *quién habla* (1), y el español el género de *qué se habla* (2).

- |     |                    |                                      |
|-----|--------------------|--------------------------------------|
| (1) | a. ikian ta irara  | ‘Esta es mi canoa’ (hablante hombre) |
|     | b. ajan tsa irara  | ‘Esta es mi canoa’ (hablante mujer)  |
| (2) | est-a es mi cano-a | (hablante hombre o mujer)            |

Tanto en español como en kukama, las distinciones de género se codifican en varias categorías, incluyendo los demostrativos, objeto de este estudio. En el español canónico, las formas femenina y masculina pueden funcionar como modificadores (3a, 4a) y formas pronominales (3b, 4b). Sin embargo, las formas neutras solo funcionan como pronombres, y generalmente hacen referencia a proposiciones completas (5). En kukama, las formas masculinas/femeninas pueden funcionar tanto como modificadores como pronombres.

- |     |  |  |
|-----|--|--|
| (3) | a. Llegó <i>esa</i> mujer a la entrevista.         |  |
|     | b. <i>Esa</i> <i>es</i> la que contacté ayer.      |  |
| (4) | a. Llegó <i>ese</i> hombre a la entrevista.        |  |
|     | b. <i>Ese</i> <i>es</i> el que contacté ayer.      |  |
| (5) | a. Tenemos <i>que</i> contactar más participantes. |  |
|     | b. <i>Eso</i> <i>es</i> lo que hice ayer.          |  |

Es importante destacar que, para usar los demostrativos como modificadores en español, los hablantes dividen su atención en dos puntos de referencia. El género del demostrativo tiene como punto de referencia el sustantivo, pero la distinción entre proximal y no proximal tiene a los interlocutores como centro deíctico. En cambio, en kukama tanto el género del demostrativo como la deixis espacial tienen al hablante como centro deíctico. El hecho de que el kukama y el español tengan diferentes perfiles tipológicos con respecto al género convierte a los demostrativos en un campo particularmente fructífero para el estudio del contacto lingüístico en estas lenguas, y más allá de ellas.

Este estudio no se centra en el uso de los demostrativos para indicar distancia espacial exofórica, sino en sus funciones endofóricas. Es decir, nos enfocamos en las relaciones entre los demostrativos y otras expresiones lingüísticas que ya han aparecido (anáfora) o aparecerán más tarde (catáfora). En el discurso, los demostrativos se utilizan principalmente para coordinar la interacción comunicativa entre los interlocutores. Los hablantes los utilizan para referirse a diversas entidades, como individuos, eventos, hechos o proposiciones. Dado que se trata de una variedad poco estudiada, este estudio explora, primero, una cuestión general, antes de pasar a preguntas más específicas:

- ¿Cuáles son los patrones de distribución global de los demostrativos en el EAP?
- ¿Cuáles son los patrones de concordancia de género entre los demostrativos y sus anáforas?
- ¿Los patrones observados sugieren algún efecto de contacto?

Para responder a estas preguntas, se analizaron entrevistas de ~50 minutos a catorce participantes. La muestra de hablantes incluyó cuatro monolingües (a los que nos referimos como MONOLINGÜES) de español de la ciudad de Iquitos, cinco bilingües español/kukama nacidos y criados en comunidades kukama (a los que nos referimos como BILINGÜES), y cinco bilingües que tienen el español como primera lengua y el kukama como lengua de herencia, también nacidos y criados en comunidades kukama (a los que nos referimos como HABLANTES DE HERENCIA). Se codificaron un total de 1416 tokens según las siguientes variables: el género del hablante; el tipo de construcción (adnominal, pronominal); el tipo de demostrativo (próximo, medio); la función (rastreo de la referencia de los participantes, presentación de un participante, deixis discursiva, deixis espacial, y deixis temporal) y la animacidad del sustantivo al que el demostrativo hace referencia (humano, animado, inanimado, abstracto).

En términos generales, la concordancia de género entre demostrativos y sus anáforas es sensible al rol sintáctico del demostrativo y a la animacidad del sustantivo. La concordancia de género también es sensible al perfil lingüístico y al género social de los participantes. Uno de los resultados más sorprendentes es la diferencia en el uso de demostrativos entre monolingües y no monolingües (bilingües y hablantes de herencia), lo cual sugiere algún efecto de contacto. En otras palabras, este hallazgo sugiere que hay ciertos indicios de influencia del sustrato, ausente en los monolingües.

## 1.2 Conceptos fundamentales

En esta sección ofrecemos las ideas fundamentales empleadas en el diseño del estudio y el análisis de los datos.

### 1.2.1 Género gramatical y género indexical

El género gramatical es muy diferente al género indexical. Según Corbett (1991, p. 4), el género gramatical es una clasificación de los nombres, y se manifiesta en patrones de concordancia. Esta clasificación se basa, por lo general,

en el género de los referentes animados (ej. el niño, la niña, el gato, la gata, etcétera), pero puede también extenderse a los no-animados, como ocurre en el español (ej. la casa, la ventana, el puente, el libro, etc.). El criterio determinante para proponer que una lengua posee género gramatical es que este se refleje en el comportamiento morfosintáctico de las palabras asociadas a los nombres (ej. artículos, demostrativos, adjetivos, etcétera). En el español canónico, el sustantivo principal determina el número y el género de todos sus determinantes y modificadores. Por consiguiente, los demostrativos reproducen los rasgos de género de los sustantivos o pronombres.

El género indexical, por el contrario, hace referencia a las distinciones que codifican el género de los participantes del acto de habla. De ahí que estos sistemas sean referidos también como habla de hombres y habla de mujeres, discurso masculino y discurso femenino, generolectos, entre otras nomenclaturas. Las áreas de la gramática en las que se manifiesta la indexicalización de género son la fonología, la morfosintaxis, y el léxico. Más allá de la gramática, también se indexicaliza el género de los hablantes en el tono de habla (*pitch*), preferencia por ciertas estructuras idiomáticas, selección de ciertas variedades, estilos, estrategias discursivas, etcétera (Lakoff, 1973; Fleming, 2012; Rose, 2015). Según Fleming (2012, p. 300), el género indexical es un fenómeno raro en las lenguas del mundo. En efecto, este autor reporta solo 17 lenguas que poseen generolectos. Rose (2015, p. 497) encontró que por lo menos 41 lenguas habladas en Sudamérica presentan indexicalidad de género en algún dominio gramatical.

Los sistemas que indexicalizan el género de los interlocutores pueden variar también en términos del participante que es indexicalizado. Es decir, las lenguas pueden potencialmente codificar el género de *quién* está hablando, el género de *con quién* se está hablando, o el género de los *interlocutores*. En función de este parámetro, Haas (1944, p. 147), propone tres tipos de sistemas:

- Tipo 1: género del hablante
- Tipo 2: género del oyente
- Tipo 3: género relacional

Las lenguas que muestran el primer tipo hacen un seguimiento al género de quien habla, mientras que las del segundo tipo lo hacen con el género del interlocutor. Las lenguas del tercer tipo codifican si el acto de habla se produce entre participantes hombres, entre mujeres, o si el grupo es mixto. Según Rose (2015, p. 499), el tipo 1 es el más común: 37 de 41 lenguas sudamericanas que poseen generolectos (90 %) son del Tipo 1. El tipo 2 es menos común; solo 3 (7 %) poseen este sistema. El tipo 3 es el más raro de todos (1/41, 3 %). Rose reporta que 4 lenguas sudamericanas poseen un sistema mixto (tipos 1 y 3). El kukama posee un sistema categórico que indexicaliza el género del hablante (tipo 1). Este sistema se manifiesta en diferentes dominios, siendo una de las lenguas con uno de los patrones de indexicalización de género más robustos dentro de los sistemas documentados hasta el momento (para una discusión más detallada del sistema kukama, ver Vallejos, 2015).

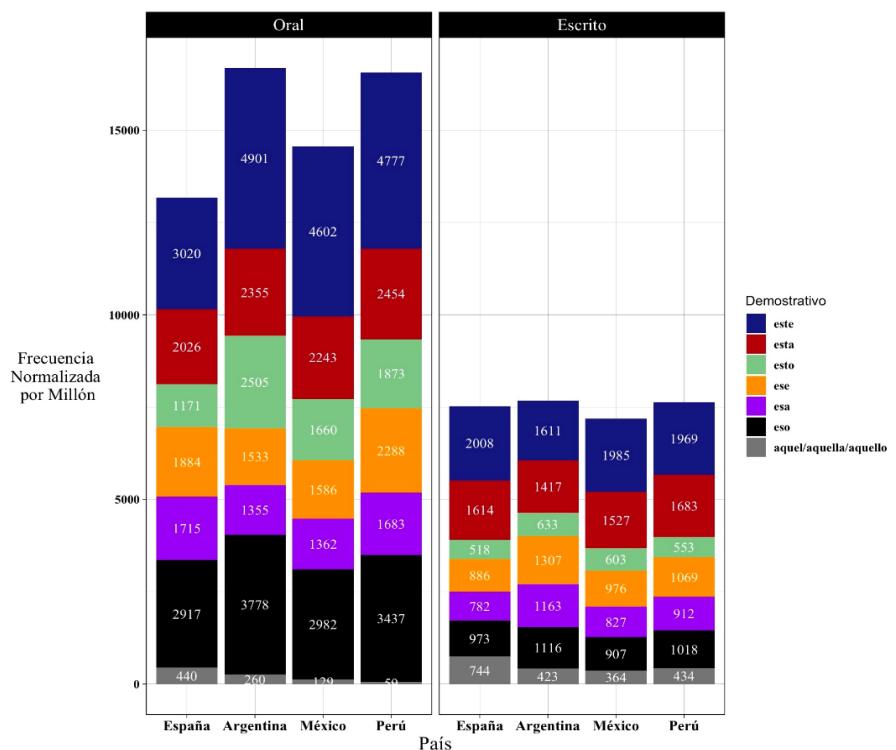
### 1.2.2 Demostrativos

Una función fundamental de los demostrativos es codificar la deixis espacial. Es decir, en su función EXOFÓRICA, se utilizan para indicar la distancia entre los hablantes y una entidad dada. Sin embargo, el énfasis de este estudio no está en los usos de los demostrativos para referirse a entidades o eventos reales o imaginarios presentes en el espacio en el momento de la enunciación. Más bien, este estudio se centra en la función ENDOFÓRICA o discursiva de los demostrativos. En el discurso, los demostrativos se utilizan principalmente para coordinar la interacción comunicativa entre interlocutores, en lugar de indicar distancia espacial. En otras palabras, los hablantes utilizan los demostrativos para referirse a una variedad de entidades como individuos, entidades, eventos o proposiciones. Los demostrativos también se emplean para establecer un foco de atención conjunto de los interlocutores (Diessel, 2006; Diesel & Coventry, 2020), para indicar el estado informativo del referente (Gundel et al., 2004) o el grado de accesibilidad del referente (Ariel, 1988).

En el español canónico, los demostrativos se agrupan en tres categorías: próximos, medios y distales. Sin embargo, la frecuencia de uso de los demostrativos canónicos del español varía según la forma, la región y la modalidad lingüística (oral versus escrita). La Figura 1, de Shin & Vallejos-Yopán (2023), presenta la frecuencia normalizada por cada millón de palabras (cpm) de cada demostrativo singular en el Corpus del Español del Siglo xxi(CORPES).

**Figura 1**

*Frecuencia de demostrativos en CORPES (Shin & Vallejos-Yopán, 2023, p. 429)*



En general, los demostrativos se usan con mayor frecuencia en datos orales que en datos escritos, con la excepción de los distales, que son más frecuentes en datos escritos. Otro punto notable es el número alto de instancias del demostrativo masculino próximo este en los datos orales. Una explicación posible es que estas cifras incluyen las funciones innovadoras de muletilla que ha desarrollado este demostrativo en Latinoamérica (Vallejos-Yopán, 2023).

La distribución de los demostrativos en la Figura 1 también muestra algunas diferencias entre países. Nótese, por ejemplo, el uso mínimo de formas distales en las variedades peruanas (59 ejemplos, 0,4 % de todos los demostrativos en datos orales de Perú). Cabe destacar que los demostrativos distales aquél, aquella, aquello se combinaron dada su baja frecuencia general. Este hallazgo ha llevado a algunos académicos a proponer que, en algunas variedades, el sistema español podría ya no ser tripartito, sino bipartito con respecto a la distancia espacial. Así ocurre en el corpus de español amazónico construido para este estudio, donde no se han documentado ejemplos de las formas distales aquella, aquél, aquello, solo las formas próximas y medias. En la variedad en estudio, las formas medias se emplean también para cubrir el ámbito funcional de las formas distales.

Este estudio aborda los demostrativos femeninos, masculinos y neutros del español. Las formas relevantes del español se resumen en la Tabla 1 y las formas kukama en la Tabla 2.

**Tabla 1**  
*Demostrativos en español amazónico peruano*

	<b>Femenino</b>	<b>Masculino</b>	<b>Neutro</b>
<b>Próximo</b>	<i>esta</i>	<i>este</i>	<i>esto</i>
<b>Medio</b>	<i>esa</i>	<i>ese</i>	<i>eso</i>

Las formas femenina y masculina pueden funcionar como modificadores en construcciones adnominales, como en (3a) y (4a), o como formas pronominales en construcciones pronominales, como en (3b) y (4b). En estos usos sintácticos, los demostrativos femeninos y masculinos reproducen los rasgos de género de sus anáforas. Los demostrativos neutros difieren de los femeninos y masculinos en algunos aspectos. En primer lugar, no participan en construcciones adnominales. Además, pueden referirse a varias nociones, pero generalmente son entidades inanimadas, sean concretas o abstractas, especialmente cuando el hablante no está seguro de su identidad (Vallejos-Yopán, 2023). Los demostrativos neutros no suelen usarse para referirse a animales o humanos, a menos que se quiera referirse a ellos de manera ofensiva. Así, mientras que la expresión “este que está aquí” puede referirse a una persona, un animal o una cosa, la expresión “esto que está aquí” se interpreta automáticamente como haciendo referencia a una entidad inanimada. Finalmente, las formas neutras se usan para hacer referencia a proposiciones completas, como se muestra en

(5). Esto será importante al codificar los demostrativos neutros respecto a la concordancia o discordancia de género, como se explica en §6.3.3.3.

**Tabla 2**  
*Demostrativos en kukama*

	<i>Mujeres</i>	<i>Hombres</i>
<i>Próximo</i>	<i>ajan</i>	<i>ikian</i>
<i>Distal</i>	<i>yukun</i>	<i>yukan</i>

Como se muestra en la Tabla 2, el kukama cuenta con un conjunto de demostrativos que también difieren en el parámetro próximo/no-próximo. Sin embargo, a diferencia del español, las formas kukama indican el género social del hablante, no del referente del que se habla. Hay dos puntos clave que deben considerarse. En primer lugar, en las interacciones espontáneas kukama, el centro deíctico se reenfoca constantemente cuando los hablantes incluyen citas directas de otros en su discurso. Por ejemplo, si un hablante masculino cita a un personaje femenino, las formas demostrativas utilizadas en la cita son las femeninas. En segundo lugar, quienquiera que esté hablando se interpreta obligatoriamente como animado. En resumen, los hablantes de kukama prestan atención a la animación de los referentes y al género de los participantes en las citas directas al usar demostrativos. Esto podría tener un efecto en la manera como emplean los demostrativos en español.

En general, existe la creencia de que los demostrativos son un área de la gramática menos sensible a los efectos de contacto. Según Ramat (2015), “en el ámbito de los demostrativos apenas encontramos casos de préstamo entre las lenguas romances” (p. 585). Si bien esto puede ser cierto, los hallazgos de este estudio muestran que, independientemente del préstamo de formas, son posibles efectos más sutiles en el uso de demostrativos.

### **1. 2. 3 *Lengua de herencia y perfil sociolinguístico***

Para fines de este estudio, es importante distinguir entre segunda LENGUA Y LENGUA DE HERENCIA. Una segunda lengua es cualquier lengua que una persona aprende además de su primera lengua, lo cual no implica ningún conocimiento previo de esta lengua. En el contexto de las comunidades indígenas amazónicas, una segunda lengua por lo general puede ser el español, el portugués, o el inglés. A diferencia de una segunda lengua, una lengua de herencia implica cierta exposición a dicha lengua durante la infancia, así como una conexión cultural con ella a través de la familia, los ancestros, el pueblo indígena o el país de origen. En esta perspectiva, una lengua de herencia puede ser una lengua indígena en proceso de desplazamiento, como el kukama y el kichwa en Perú (véase Vallejos-Yopán et al., 2022) o una lengua inmigrante (como el español en los EE. UU. o el turco en Alemania).

Existen definiciones restrictivas y definiciones más inclusivas de los hablantes de herencia. Según una visión restrictiva, los hablantes de herencia son considerados bilingües aun cuando la lengua que aprendieron en la infancia

ya no es su lengua dominante. Estos hablantes de herencia suelen comunicarse en la lengua mayoritaria de la sociedad en la que viven. Según una definición más inclusiva, los hablantes de herencia son quienes han tenido por lo menos alguna exposición a esta lengua durante su infancia, aun cuando no crecieron usándola como medio de comunicación (Polinsky & Kagan, 2007; Rothman, 2009). Por lo tanto, los hablantes de herencia pueden mostrar niveles de competencia comunicativa muy diversos. En este estudio adoptamos una definición amplia de hablantes de herencia. En el contexto kukama, los hablantes de herencia han sido expuestos a esta lengua en sus comunidades al crecer rodeados de sus abuelos y adultos mayores. Tienen cierto conocimiento de kukama, pero se comunican cotidianamente en español.

El que los hablantes de herencia hayan adquirido conocimientos lingüísticos en la infancia ha abierto importantes caminos en la investigación sobre la adquisición de lenguas, en particular, el rol de la edad, la calidad del *input*, y el conocimiento implícito frente al conocimiento explícito (Valdés, 2005). Por ejemplo, los hablantes de herencia parecen tener una ventaja sobre los aprendices de L2 en cuanto a fonología, y también en algunos aspectos de la morfología y sintaxis (Montrul, 2008). Las cuestiones relacionadas con los tipos de patrones que surgen en contextos de desplazamiento y revitalización de lenguas (Hinton, 2001) también constituyen importantes cuestiones teóricas para el cambio y el contacto lingüístico. El presente estudio asume esa perspectiva al considerar dos tipos de bilingües, aquellos con español como segunda lengua, y aquellos con kukama como lengua de herencia.

### 1.3 Metodología

Aunque el trabajo sobre el español amazónico es limitado, se han descrito algunas características de esta variedad. Además de un estudio sobre la evolución de *este* de un demostrativo a una muletilla (Vallejos-Yopán 2023), hasta donde sabemos no existen estudios sobre el uso discursivo de los demostrativos en esta variedad. En otras palabras, no hay investigaciones que podamos tomar como punto de referencia o metodología que podamos replicar.

#### 1.3.1 Participantes

Para explorar las preguntas de investigación planteadas anteriormente, examinamos el habla de catorce hablantes de EAP. Cinco nacieron y crecieron en la ciudad de Iquitos, y nueve nacieron y crecieron en pueblos kukama. Los datos provienen del corpus de español amazónico de Vallejos-Yopán (2023), que actualmente consta de unas 30 horas de entrevistas espontáneas recopiladas con un total de 44 hablantes. Las entrevistas examinadas en este estudio se realizaron entre 2011 y 2016 en las comunidades de habla relevantes: las comunidades kukama y la ciudad de Iquitos. La duración total de las entrevistas oscila entre 50 y 65 minutos. La entrevistadora introdujo el primer tema de conversación (por ejemplo, los eventos principales, su experiencia en la escuela secundaria/universidad, actividades preferidas, etcétera), tras lo cual los participantes tomaron la iniciativa en su mayor parte. Para recopilar datos y conversacionales de calidad en el entorno más relajado posible, el sonido se grabó con un micrófono de solapa omnidiireccional de condensador (AT80). Las entrevistas se

transcribieron en unidades de entonación, segmentos de habla definidas por características prosódicas como el tono, un patrón de entonación distintivo, pausas, y la mayoría presenta un único pico de acentuación (Chafe, 1994; Him-melmann et al., 2018).

La muestra del presente estudio está equilibrada en cuanto a perfil sociolingüístico (5 monolingües, 5 hablantes de herencia y 4 bilingües) y género de los participantes (8 mujeres y 6 hombres). Sin embargo, fue imposible lograr un equilibrio respecto al lugar de origen, debido a que tanto los bilingües como los hablantes de herencia viven en comunidades indígenas. También intentamos incluir diversidad de ocupaciones para los monolingües de Iquitos y los hablantes de herencia que viven en las comunidades. Sin embargo, no fue posible hacer lo mismo para los bilingües, debido a que el *kukama* es una lengua en grave peligro de extinción, hablada actualmente solo por ancianos.

En la sociedad *kukama*, los ancianos son una categoría social en sí misma; son modelos a seguir y se los tiene en alta estima por su sabiduría y experiencias de vida. Por lo tanto, los participantes bilingües no difieren en edad (todos son mayores a 60 años) u ocupación. Para controlar la edad en los demás participantes, seleccionamos hablantes monolingües entre los 35 a 41 años, y hablantes de herencia entre los 21 a 45 años. En cuanto a la educación formal, la mayoría de los adultos mayores solo han cursado los primeros grados de la primaria, mientras que los hablantes más jóvenes, tanto de la comunidad como de la ciudad, han tenido acceso a más de 12 años de escolarización. La Tabla 3, a continuación, muestra los datos demográficos de los hablantes.

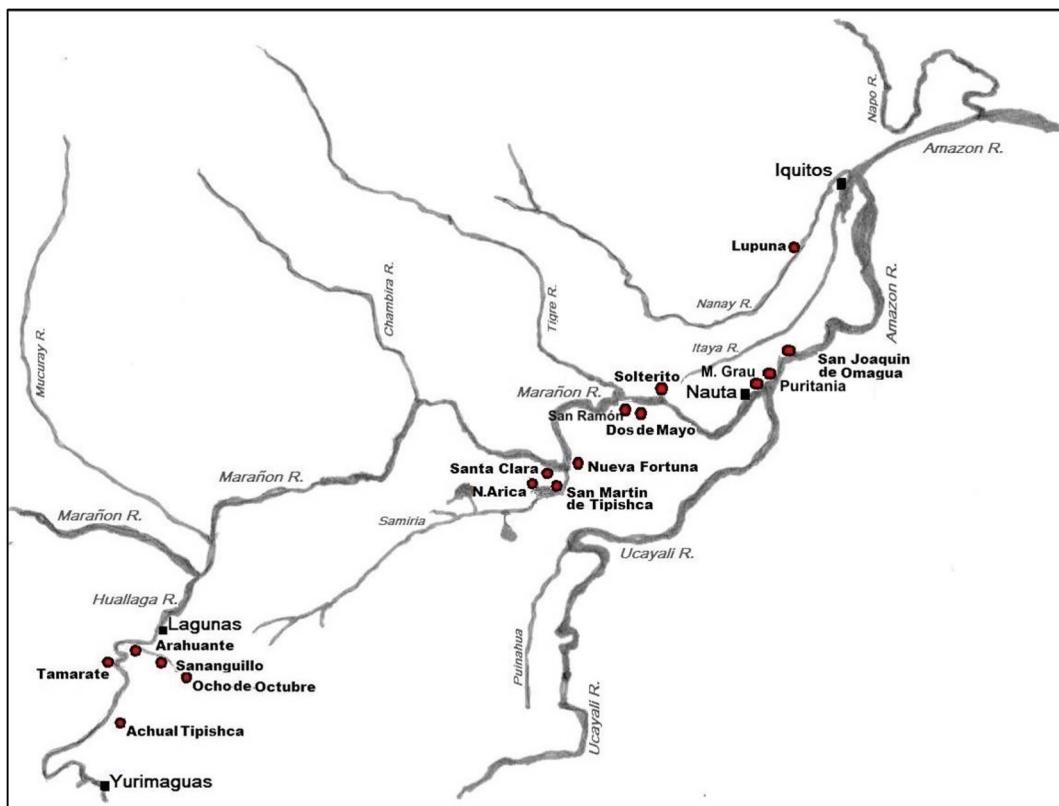
**Tabla 3**  
*Participantes*

<i>Lugar de origen</i>	<i>Participante</i>	<i>Perfil sociolingüístico</i>	<i>Género</i>	<i>Edad</i>	<i>Ocupación</i>
<i>kukama</i>	P1	bilingüe	F	64	abuela
	P2	bilingüe	F	63	abuela
	P3	bilingüe	M	68	abuelo
	P4	bilingüe	M	62	abuelo
	P5	herencia	F	45	negociante
	P6	herencia	F	32	profesora
	P7	herencia	F	21	estudiante
	P8	herencia	M	35	agricultor
	P9	herencia	M	21	estudiante
<i>Iquitos</i>	P10	monolingüe	F	37	enfermera
	P11	monolingüe	F	39	profesora
	P12	monolingüe	F	35	estilista
	P13	monolingüe	M	39	profesor
	P14	monolingüe	M	41	trabajador público

El área donde viven los participantes se indica en el Mapa 1. En las siguientes secciones ofrecemos más detalles de cada grupo según su perfil sociolíngüístico.

### Mapa 1

*Comunidades de origen de los participantes*



#### 1.3.1.1 *Bilingües con kukama como L1 y español como L2*

Los participantes incluidos en este grupo declararon tener el kukama como L1 y el español como L2. Estos participantes bilingües provienen de cuatro comunidades kukama, todas indicadas en el Mapa 1: San Martín de Tipishca (río Samiria), Dos de Mayo (río Marañón), Miguel Grau (río Amazonas) y Ocho de Octubre (río Huallaga). La distancia entre estas comunidades e Iquitos varía de 18 a 40 horas en bote. Estas comunidades han experimentado el desplazamiento del kukama en favor del español desde hace aproximadamente seis décadas. Como mencionamos anteriormente, actualmente los hablantes de kukama son escasos, y ninguno es monolingüe (Vallejos, 2016).

Al momento de las visitas de campo, las comunidades kukama de este estudio tenían acceso a la radio, pero no a la televisión ni al internet. En todas estas comunidades se ofrece educación primaria (K-6), pero solo San Martín de Tipishca y Miguel Grau tienen educación secundaria (7-11). Los participantes de los otros pueblos asistieron a la escuela secundaria en un pueblo kukama vecino.

### 1.3.1.2 *Hablantes de español como L1 y kukama como lengua de herencia*

El segundo grupo incluido en este estudio también proviene de comunidades kukama indicadas en el mapa: Puritania (río Amazonas), Dos de Mayo, San Ramón y Solterito (río Marañón). Este grupo tiene el kukama como lengua de herencia; es decir, ha tenido cierta exposición a esta lengua durante su infancia. Todos manifiestan una pertenencia al pueblo kukama y tienen una conexión profunda con su cultura construida a través de su familia, sus ancestros, y su región. Esta conexión es fundamental para el aprendizaje de lenguas de herencia, especialmente en contextos lingüísticos en peligro de extinción, como es el caso del kukama.

Sin embargo, algunos hablantes de herencia no se declaran bilingües, a pesar de tener cierto conocimiento del kukama. En este pueblo indígena, esa categoría parece ser un honor reservado para los ancianos que hablan el kukama como primera lengua (ver §6.2.3). Los kukamas de esta generación viajan regularmente desde las comunidades situadas a lo largo de ríos navegables hasta los centros urbanos más cercanos, como Nauta e Iquitos, por lo que interactúan regularmente con los hablantes de español de los alrededores.

### 1.3.1.3 *Monolingües en español*

Los hablantes monolingües provienen de Iquitos, el puerto más importante del río Amazonas dentro del Perú. Esta ciudad, la más grande de la Amazonía peruana, tiene una población total de 146 853 habitantes (INEI, 2017), o de 413 556 (INEI, 2017, Tomo I) si se incluye a los habitantes del área metropolitana, que incluye distritos como Belén, Punchana y San Juan. El 99 % de la población de Iquitos se declara hispanohablante (INEI, 2017). Aunque la ciudad atrae a migrantes de pueblos rurales, solo el 0,2 % de los residentes declaró saber otro idioma local, mientras que el 13,72 % eligió no responder o declaró no saber en qué idioma aprendió a hablar (INEI, 2017). Por lo general, los jóvenes de pueblos rurales van a Iquitos para trabajar y estudiar. Como los kukamas hablan español, es relativamente fácil para ellos integrarse en la vida de Iquitos en comparación con los miembros de otros grupos étnicos.

### 1.3.2 *Corpus*

El corpus de este estudio tiene un número comparable de palabras por grupo sociolingüístico. La Tabla 4 muestra la distribución de las 125 144 palabras del corpus para cada grupo.

**Tabla 4**  
*Número de palabras en el corpus*

<i>Grupo</i>	<i># palabras</i>
Bilingües	38 036
Hablantes de herencia	44 436
Monolingües	42 672

Es notable mencionar que, a pesar de que la muestra de los tres grupos es de un tamaño comparable, los hablantes bilingües produjeron una cantidad mayor de demostrativos, como se mostrará más adelante en §6.4.2.5. Adicionalmente, observamos ciertas colocaciones frecuentes en la base de datos, como *por eso* y *a raíz de eso*. Estas secuencias se incluyeron en este estudio, ya que en esta variedad es posible reemplazar *eso* por otra forma neutra, como en *por esto* y *a raíz de esto*. Por otra parte, excluimos 33 instancias del análisis. Estas incluyen oraciones fragmentadas o incompletas, interrupciones, y vacilaciones, como en (6a), demostrativos interrogativos, como en (6c), y casos donde no logramos identificar la anáfora del sustantivo o resultó ser ambigua, como explicaremos en §6.3.3.3. Las repeticiones, como en (6b), se incluyeron solo una vez. En los ejemplos, la separación entre unidades entonacionales está indicada con una barra inclinada (/).

- (6) a. Y eso que... / eso...  
b. Y ese / ese carro...  
c. ¿Qué es eso?

Finalmente, hubo 422 instancias de *este* y *ese* con uso de muletilla. Cuando ocupa esta función, *este* es semánticamente independiente del discurso, como se muestra en los ejemplos de (7).

- (7) a. ella me dice este / desaparecieron todos  
b. para comprar, este / leche caspi y caucho  
c. el señor, este / Manuel creo que se llama este señor  
d. acá había fábrica de botones antes, entonces / este / ahí  
e. Yo después ya / de años me / ese yo también estudiando así

Estos fragmentos poseen propiedades fonológicas y prosódicas particulares, e incluso su propia distribución sintáctica. Por lo tanto, excluimos de este estudio las formas *este* y *ese* con función de muletilla (para un análisis de la forma *este* en función de muletilla, ver Vallejos-Yopán, 2023). Después de esta primera exclusión, el número total de instancias que se analizó para este estudio fue de 1416.

### 1.3.3 *Variables*

Teniendo en cuenta la literatura introducida en 6.2, los 1416 demostrativos identificados en el corpus fueron codificados en cuanto a variables extralingüísticas y variables lingüísticas. Para las variables extralingüísticas, cada instancia fue codificada según el género social del hablante, su lugar de residencia (comunidad kukama/ciudad de Iquitos), y su perfil sociolingüístico (monolingüe, bilingüe o hablante de herencia). Es importante mencionar que en este estudio no incluimos la edad de los participantes porque nuestra muestra no es balanceada en este aspecto, como se presentó en 6.3.1. A continuación, explicamos cómo hemos codificado cada variable lingüística.

### 1.3.3.1 Función sintáctica

Cómo se explicó en la introducción, los demostrativos femeninos y masculinos pueden aparecer en función adnominal (ver ejemplos 3a, 4a) o pronominal (ver ejemplos 3b, 4b). Las formas neutras solo funcionan como pronombres, como en (5). En (8) y (9) ilustramos la codificación de esta variable.

#### Adnominal

- (8) a. *Esos* perros de la calle traen la peste pues  
 b. ellos han, han fundado *ese* pueblito.

#### Pronominal

- (9) a. *Esa* es la que contacté ayer.  
 b. El hombre me seguía pues, ahí *ese* todavía no era mi marido

### 1.3.3.2 Género del núcleo nominal y de la anáfora

La codificación de los sustantivos fue necesaria para establecer la concordancia de género entre el demostrativo y su núcleo nominal (en construcciones adnominales), o su anáfora (en construcciones pronominales). Identificamos dos categorías de sustantivos que ilustramos en (10) y (11).

#### Femenino

- (10) a. O sea que esta *casa* que tengo ahora.  
 b. El mono, por más que vea *candela*, eso no lo ahuyenta.

#### Masculino

- (11) a. Esos *perros* de la calle traen la peste pues  
 b. ellos han, han fundado *ese* *pueblito*.

### 1.3.3.3 Concordancia de género

Este factor se deriva del género del demostrativo mismo y el género del sustantivo (ver §6.3.3.2) y tiene dos niveles: concordancia y discordancia. Las formas femeninas (*esa*, *esta*) y masculinas (*ese*, *este*) se codificaron como concordancia si su anáfora era un sustantivo femenino o masculino, respectivamente. Véanse los ejemplos en (12). Los demás escenarios se codificaron como discordancia. Esto incluye los frecuentes casos en el corpus donde las formas neutras (*eso*, *esto*) hacen referencia a sustantivos masculinos o femeninos, incluyendo animales y seres humanos, como se muestra en (13).

#### Concordancia

- (12) a. O sea que *esta casa* que tengo ahora.  
 b. El hombre me seguía pues, ahí *ese* todavía no era mi marido  
 c. Me necesitarán ahí, pienso, pero *eso* no es seguro

### Discordancia

- (13) a. Hizo *ese aula*, nada más, y deja ahí nomás.  
b. Ese [piscina] está así nomás  
c. Criábamos gallina en la chacra, y *eso* vendíamos para la fiesta.

Los casos ambiguos, aquellos en los que la anáfora podía interpretarse como más de una entidad, e incluso una proposición, o donde nos fue imposible identificar la anáfora fueron excluidos del análisis, como mencionamos anteriormente en §6.3.2.

#### 1.3.3.4 Animacidad

Para este estudio, adaptamos la jerarquía de animacidad de Zaenen et al. (2004). Esta jerarquía consta de diez categorías: humanos, organizaciones, animales, autómatas (computadoras, robots), vehículos, lugares, tiempos, concretos, no concretos y mixtos. Las categorías relevantes para explicar las instancias documentadas en el corpus del EAP son cuatro: humanos (14a), animados (animales y plantas) (14b), inanimados (concretos) (14c) y abstractos (no concretos) (14d). Requerimos de una categoría adicional, proposiciones, para cubrir todas las instancias documentadas en el corpus (14e). En los ejemplos a continuación, el demostrativo objetivo está en negrita y la anáfora está subrayada.

- (14) a. La muchacha estaba llorando, *esa* era su hija de mi tía  
b. *Esos perros* de la calle traen la peste pues  
c. Osea no es *esta casa* que tengo ahora  
d. Nos recibió con *esa emoción*  
e. Vio todo y siguió como si nada, *eso* no le importaba

#### 1.3.3.5 Funciones discursivas

Identificamos tres funciones principales de los demostrativos en la base de datos: endófora o rastreo de referentes, presentación de participantes, y deixis (Lyons, 1977; Himmelmann, 1996; Diessel, 2006). En el rastreo de referentes, los demostrativos hacen referencia a los participantes del discurso. Sirven para hacer el seguimiento a los referentes (en inglés, *reference tracking*), para establecer contraste con otro referente o cambiar el foco de atención. Existen dos tipos de endófora: ANÁFORA, cuando el referente aparece antes, y CATÁFORA, cuando el referente aparece después. La gran mayoría de instancias en el corpus corresponden a la función anafórica, aunque no hemos codificado estos subtipos para fines de este estudio. Esta función se ilustra en (15). Como vemos en (15a), el demostrativo *eso* se refiere a la *candela*, mientras que en (15c) vemos una situación paralela, donde *esos* se refiere a los *trabajadores* mencionados anteriormente.

- (15) a. El mono, por más que vea candela, *eso* no lo ahuyenta.  
b. Grau es todito kukama / [...] / ellos han, han fundado *ese* pueblito.

- c. Los trabajadores se molestaron / dice no les pagaban / nada les daban / entonces *esos* trabajadores hablaron con sus jefe.

Para propósitos de este estudio, identificamos dos tipos de presentación: la identificación de un participante y la incorporación de un nuevo participante en el discurso. La identificación de un participante se realiza a través de construcciones técticas en las que el demostrativo es el sujeto de una construcción con cópula. Esta función se ilustra en (16).

- (16) a. *Este* es el putsui, que le decimos también huanchaco  
 b. *Esta* es [la comunidad] Tangarana  
 c. *Ese* es una patarashca

Por otra parte, la introducción de un nuevo participante en el discurso se refiere a situaciones en las que un personaje se menciona por primera vez en el discurso. En el corpus, esta función aparece en construcciones adnominales, al contrario de la identificación, donde la construcción es pronominal. Esta función se ilustra en (17). En (17a) vemos también un demostrativo medio innovador en el EAP para el masculino plural, *eses*.

- (17) a. *Esos* dos candidatos tenían buena / propuesta.  
 b. Es igual que ha hecho *ese* / Salomón Abensur.

Por último, hemos codificado tres tipos de **deixis** que involucran a demostrativos: deixis discursiva, deixis temporal y deixis espacial. La deixis discursiva remite a proposiciones, segmentos adyacentes en el discurso. En otras palabras, remite a fragmentos que ya están presentes en el entorno comunicativo, y que, por lo tanto, son susceptibles a ser señalados. Los ejemplos en (18) ilustran esta función.

- (18) a. Ajá, *eso* era el primer requisito / hablar la lengua  
 b. Vio todo y siguió como si nada / *eso* no le importaba

La deixis temporal hace referencia a un momento en el tiempo. Se expresa de forma más frecuente a través de adverbios temporales (como *hoy*, *ayer*, *el mes próximo*) o expresiones que contienen demostrativos, como *ese/este momento*, *esa/esta noche*, entre otras. Los ejemplos en (19) ejemplifican esta función.

- (19) a. Destrozó su asiento, y en *ese* momento llega la vieja  
 b. Ya, y *ese* día / le hemos llevado a la posta

Finalmente, la deixis espacial hace referencia a un lugar en el espacio donde sucede un evento. Se expresa mediante adverbios locativos como *ahí*, *allá*, etc. Algunas combinaciones locativas frecuentes que contienen demostrativos son *en ese/este lado*, *esa/esta calle*, *ese/este lugar*, etcétera. Ejemplos del corpus se ofrecen en (20).

- (20) a. ahí en *ese* camino por donde sale al maizal / ahí se le hace una barbacoa
- b. Vivo en la comunidad de Payorote / [...] / también viven en *ese* lugar mis abuelitos

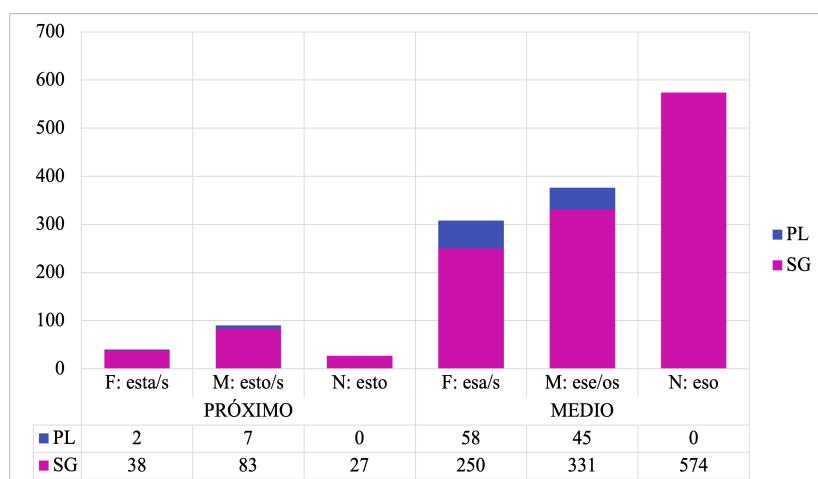
## 1.4 Resultados

A continuación, presentamos los resultados más importantes del análisis cuantitativo. Comenzamos con el uso global de los demostrativos en el corpus, para luego proceder a discutir cada una de las variables significativas y su peso relativo en la discordancia de género.

### 1.4.1 Uso global de demostrativos

En total, identificamos 1416 demostrativos en el corpus. Esta cifra excluye 422 instancias de *este* y *ese* como muletillas (*fillers*, en inglés). La Figura 2 ofrece la distribución global de los 1416 demostrativos organizados según su forma.

**Figura 2**  
*Distribución global de los demostrativos*



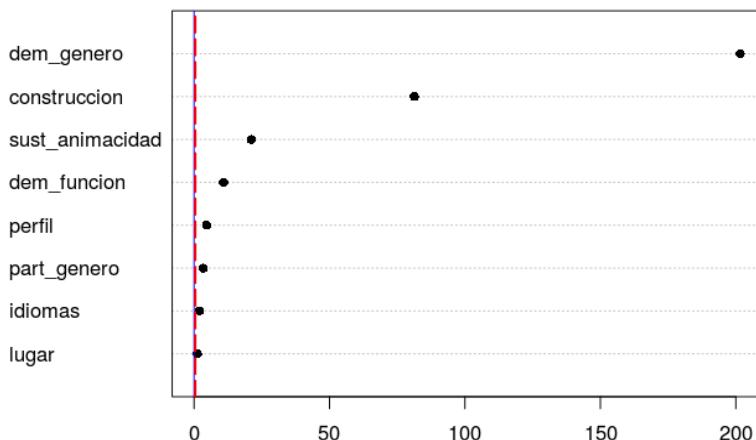
Como se puede observar en la Figura 2, identificamos 157 instancias de demostrativos próximos en el corpus, una cifra mucho menor que la de demostrativos medios, 1258. El demostrativo neutro *eso* tiene un uso particularmente alto, con 574 instancias. La alta frecuencia de *eso* es una primera señal de su “sobreuso” en la variante en estudio. Esto sugiere que, además de hacer referencia a proposiciones, el demostrativo *eso* abarca espacios funcionales en el EAP que son asumidos por formas femeninas o masculinas en variedades canónicas del español.

### 1.4.2 Variables significativas

Como el objetivo de este artículo es identificar los escenarios en los que observamos discordancia de género entre el demostrativo y su núcleo nominal, en las siguientes secciones presentamos las variables que tienen un rol significativo en la concordancia de género. Para este análisis, hemos excluido 313 instancias donde la anáfora es una proposición, además de las formas con menos de 5

instancias en el corpus, *estes* y *estas*. También excluimos todas las instancias de una participante que tuvo respuestas categóricas (donde ningún uso mostró discordancia). Después de todas las exclusiones, las instancias incluidas en el análisis estadístico de esta sección son 1044, provenientes de 13 participantes. Los resultados de la evaluación de los datos con Random Forest (Breiman, 2001) se ofrecen en la Figura 3.

**Figura 3**  
*Variables significativas en la discordancia de género según Random Forest*

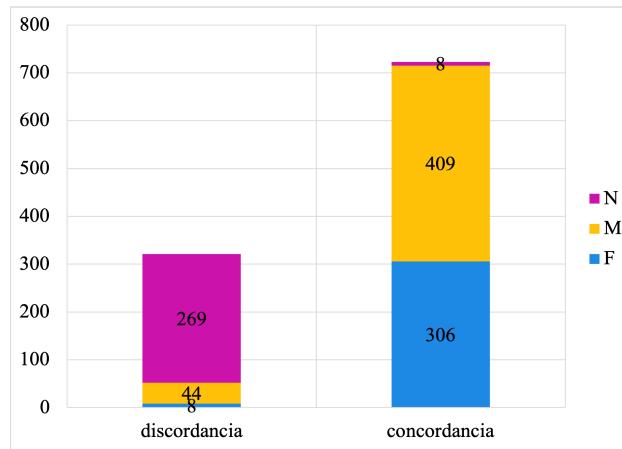


Como muestra la Figura 3, las variables que juegan un rol en los patrones de concordancia son fundamentalmente las lingüísticas. En primer lugar está el género del demostrativo (dem\_genero), seguido por la función sintáctica del demostrativo (construccion), la animacidad del sustantivo (sust\_animacidad), y la función del demostrativo (dem\_funcion). El perfil sociolingüístico de los participantes (perfil) y el género social del participante (part\_genero) también juegan un rol, aunque es menor comparado con el de las variables lingüísticas. Por último, el manejo de idiomas, es decir, el hecho de que un hablante identifique su L1 como español o kukama, no tuvo un impacto, como tampoco el lugar de origen. Volveremos a este punto, que consideramos uno de los hallazgos más importante de este estudio, en §6.5.

#### 1.4.2.1 *Género del demostrativo y concordancia de género*

Como se puede observar en la Figura 3, la variable más importante para predecir la discordancia de género en el EAP es el género del demostrativo, es decir, si es masculino, femenino o neutro. Como vemos en la Figura 4, los demostrativos que están directamente asociados con la discordancia son los neutros, lo cual es confirmado por el análisis de regresión (ver §6.4.3).

**Figura 4**  
*Concordancia de género según el género del demostrativo*



De los 277 casos con demostrativos neutros, el 42 % hace referencia a entidades femeninas, mientras que el 58 % hace referencia a entidades masculinas. Por otra parte, todos los casos de discordancia donde el demostrativo es masculino o femenino hacen referencia a sustantivos pertenecientes al otro género.

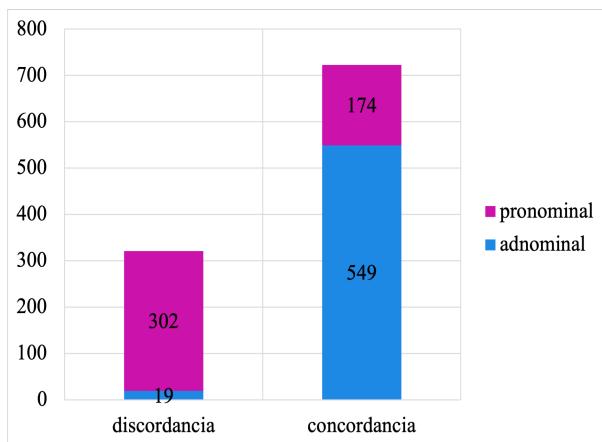
#### 1.4.2.2 Función sintáctica del demostrativo y concordancia de género

La segunda variable en importancia es la función sintáctica del demostrativo. Según estudios tipológicos, el uso adnominal tiende a ser claramente más frecuente que el pronominal en el discurso (Himmelmann, 1996, p. 214). En el EAP vemos que, aunque esta tendencia se mantiene, la distribución de las dos funciones es relativamente similar. Sin embargo, como se muestra en la Figura 5, esta variable ayudará a entender si la concordancia o discordancia de género está supeditada a la función sintáctica del demostrativo.

**Tabla 5**  
*Demostrativos según su función sintáctica*

Función sintáctica	Instancias	%
Adnominal	568	54
Pronominal	476	46

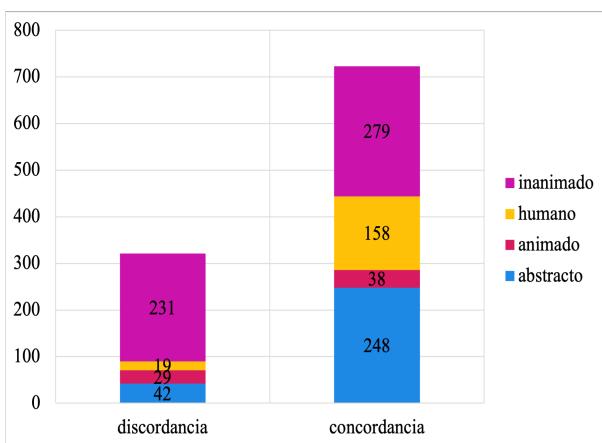
Como indica la Tabla 5, los casos donde la función sintáctica del demostrativo es pronominal son una minoría en el total de los datos (46 %). Sin embargo, si se mira solamente a las instancias donde existe discordancia de género, los demostrativos pronominales constituyen la gran mayoría de casos (94 %).

**Figura 5***Concordancia de género según la función sintáctica del demostrativo***1.4.2.3 Animación del sustantivo y concordancia de género**

La tercera variable en importancia respecto a la concordancia de género es la animación del sustantivo. Casi la mitad de sustantivos (49 %) son inanimados, seguido por abstractos, humanos, y animados. La frecuencia de cada categoría de animación se ofrece en la Tabla 6.

**Tabla 6***Sustantivos según su categoría de animación*

Animación	Instancias	%
Inanimado	510	49
Abstracto	290	28
Humano	177	17
Animado	67	6

**Figura 6***Concordancia de género según la animación del sustantivo*

La Figura 6 muestra que los sustantivos inanimados, que componen el 49 % del total de sustantivos, son aún más frecuentes cuando examinamos solamente la categoría de discordancia de género (72 %), mientras que tan solo componen el 39 % de la categoría de concordancia.

#### 1.4.2.4 Función discursiva del demostrativo y concordancia de género

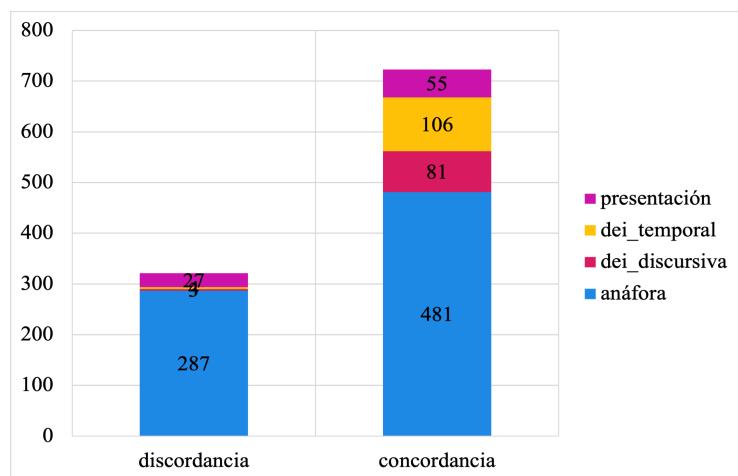
La distribución de los demostrativos según su función discursiva se ofrece en la Tabla 7. En el corpus para este estudio, la gran mayoría de los demostrativos cumple una función de anáfora. El resto de funciones ocurren con una frecuencia mucho menor de entre 8 y 11 por ciento<sup>1</sup>.

**Tabla 7**  
*Demostrativos según su función discursiva*

Función discursiva	Instancias	%
Anáfora	768	74
Deixis temporal	110	11
Deixis discursiva	84	8
Presentación	82	8

La Tabla 7 muestra que los demostrativos con función anafórica son el 74 % de los casos totales. Este porcentaje aumenta al 89 % cuando evaluamos la interacción de la función anafórica y la discordancia de género. La Figura 7 a continuación indica la distribución de la concordancia de género según la función del demostrativo.

**Figura 7**  
*Concordancia de género según la función discursiva del demostrativo*



<sup>1</sup> En el corpus encontramos 41 casos de deixis espacial, pero estas no involucraban formas próximas. Por esta razón, esas 41 instancias fueron excluidas del análisis de regresión.

#### 1.4.2.5 Concordancia según el perfil sociolíngüístico de los hablantes

Otro factor a considerar respecto a la concordancia de género es el perfil sociolíngüístico de los hablantes; es decir, si son bilingües kukama-español, hablantes kukama de herencia cuya primera lengua es el español, o monolingües en EAP, como se describió en §6.3.1. La Tabla 8 ofrece la distribución de los demostrativos en función al perfil sociolíngüístico de los hablantes.

**Tabla 8**

*Instancias según el perfil sociolíngüístico de los participantes*

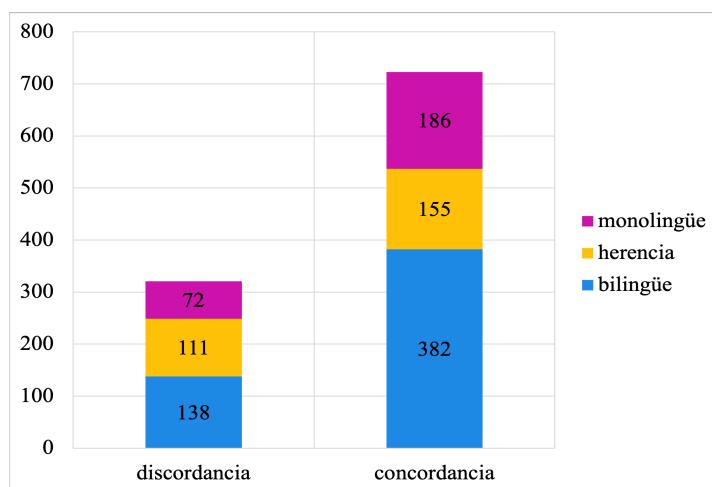
Función discursiva	Instancias	%
Bilingüe	520	50
Herencia	266	25
Monolingüe	258	25
Total	1044	100

Como muestra la Tabla 8, aproximadamente la mitad del corpus proviene del grupo de hablantes bilingües. El número de instancias para los otros dos grupos es muy similar: cada uno representa el 25 % de demostrativos del corpus.

Cuando examinamos los casos de discordancia en la Figura 8, los monolingües representan el 22 %, los hablantes de herencia el 35 %, y los bilingües el 43 %. En el caso de la concordancia, los monolingües representan el 26 %, los hablantes de herencia 21 %, y los bilingües el 53 %. Un punto notable es que, mientras que el porcentaje de los casos de concordancia es mayor al de los casos de discordancia en cuanto a los monolingües y bilingües, lo opuesto ocurre para los hablantes de herencia.

**Figura 8**

*Concordancia de género según el perfil sociolíngüístico de los hablantes*



### 1.4.3 Peso de las variables

Para examinar con mayor precisión el peso de cada variable discutida en §6.4.2, realizamos un análisis de regresión lógistica binomial (Bates et al., 2015)<sup>2</sup>. Este arrojó resultados similares a los del método de Random Forest. Las variables independientes introducidas en el modelo son el género del demostrativo (F, M, N), la función sintáctica del demostrativo (adnominal, pronominal), la animacidad del sustantivo (humano, animado, inanimado, abstracto), y la función del demostrativo (anáfora, presentación de participantes, deixis discursiva/temporal/espacial). Por otro lado, incluimos también el perfil sociolingüístico de los participantes (monolingüe, bilingüe, hablante de herencia) y el género del participante.

Los resultados del modelo que da mejor cuenta de los datos se presentan en la Tabla 9 (resultados de R Studio). Dado que el modelo se configuró para predecir la discordancia, la variable dependiente fue una variable binaria compuesta por los valores sí/no.

**Tabla 9**

*Variables significativas en el análisis de regresión (n=1044)*

Factor	Estimado	Error estándar	Valor-Z	Pr(> z )
Género del demostrativo: Neutro	6,92097	0,66331	10,434	< 2e-16***
Género del demostrativo: Masculino	0,92855	0,44039	2,108	0,03499*
Construcción: Pronominal	2,05562	0,39376	5,220	1,79e-07***
Animacidad del sustantivo: Humano	-1,63291	0,73436	-2,224	0,02618*
Animacidad del sustantivo: Inanimado	1,15064	0,52172	2,205	0,02742*
Género del participante: Masculino	-1,00449	0,33478	-3,000	0,00270**
Perfil sociolingüístico: Monolingüe	-1,65743	0,51721	-3,205	0,00135**

Un coeficiente positivo indica que la variable aumenta la probabilidad de la discordancia de género, mientras que un coeficiente negativo indica que la variable aumenta la probabilidad de la concordancia de género. Las anotaciones a la derecha indican el valor p: 0 ‘\*\*\*’, 0,001 ‘\*\*’, 0,01 ‘\*’, 0,05 ‘.’, y 0,1 ‘’.

El modelo revela que la discordancia tiene una probabilidad significativamente mayor de ocurrir cuando el género del demostrativo es neutro (*eso, esto*) y en menor medida, masculino (*ese, esos, estos*). Otro escenario que promueve la discordancia es cuando el demostrativo tiene una función pronominal. Respecto a la variable animacidad del sustantivo al que el demostrativo hace referencia, los sustantivos inanimados promueven la discordancia, mientras que los humanos promueven la concordancia. Por último, en cuanto a las variables extralingüísticas, los participantes de género social masculino y los monolingües en español promueven la concordancia de género. La función discursiva del demostrativo y el idioma L1 de los participantes no resultaron ser variables significativas.

<sup>2</sup> Un modelo mixto con hablantes como efecto aleatorio dio los mismos resultados.

## 1.5 Discusión

A continuación, discutimos los hallazgos presentados en la sección de resultados organizados en torno a las preguntas de discusión planteadas en la introducción.

### 1.5.1 *¿Cuáles son los patrones de distribución global de los demostrativos en el EAP?*

Como primera observación, los demostrativos próximos (11 %) son mucho menos frecuentes que los medios (89 %) en el uso global, mientras que los distales están completamente ausentes. Esto indica una realineación de los demostrativos en esta variante de un sistema tripartito a uno bipartito. Esta diferencia entre el porcentaje de los demostrativos próximos y medios en el EAP se puede explicar en parte por el demostrativo medio neutro *eso*, cuyo uso fue el mayor de todos los demostrativos (46 %). Esta alta frecuencia parece ser un rasgo único de esta variante; si se compara con los datos sobre la distribución de demostrativos en variantes orales del español en CORPES, presentados en Shin y Vallejos-Yopán (2023), es mucho mayor que la frecuencia de *eso* en España (22 %), Argentina (23 %), México (20 %) y Perú (21 %).

Dado que la meta es explicar la concordancia de género, para la mayoría del análisis, excluimos también los casos donde los demostrativos tienen como referente a una proposición y nos enfocamos solamente en los que se refieren a entidades a través de sustantivos. Tomando en cuenta esta segunda exclusión, el 43 % de los demostrativos son masculinos, el 31 % son femeninos, y el 25 % son neutros. Debido a que en la gran mayoría de casos los sustantivos del español solamente pueden ser masculinos o femeninos, esto implica que los demostrativos neutros están representando referentes masculinos o femeninos.

### 1.5.2 *¿Cuáles son los patrones de concordancia de género entre los demostrativos y sus anáforas?*

Un porcentaje significativo del total de demostrativos (31 %) exhibe discordancia de género gramatical. Cuando solamente se observa los casos que exhiben discordancia, la gran mayoría (84 %) son neutros, mientras que el 14 % son masculinos, y el 2 % son femeninos. Por otra parte, entre los demostrativos que exhiben concordancia de género (69 % del total), 42 % son femeninos, 57 % son masculinos y solamente 1 % son neutros. Esto se alinea con lo que muestra el análisis de regresión; la selección de un demostrativo neutro es el factor más significativo respecto a la discordancia de género. Los demostrativos masculinos también son un factor relevante, aunque en menor medida.

Adicionalmente, identificamos dos usos sintácticos principales para los demostrativos: adnominal (54 %) y pronominal (46 %). Mientras que los demostrativos están junto a sus referentes en construcciones adnominales, puede haber una distancia considerable entre el demostrativo y su referente en construcciones pronominales. Este último contexto favorece la discordancia, lo que se confirma con el análisis de regresión; este muestra que la construcción pronominal tiene un rol significativo en la discordancia de género.

Respecto a la clasificación por animación, casi la mitad de los demostrativos hacen referencia a sustantivos inanimados (49 %), seguido por abstrac-

tos (29 %), humanos (17 %), y animados (6 %). El análisis de regresión muestra una relación entre la concordancia de género y la jerarquía de animacidad; los humanos (en la cima de la jerarquía) llevan una mayor concordancia de género, mientras que los inanimados (en el fondo de la jerarquía) tienen una mayor discordancia de género.

En cuanto a la función de los demostrativos, predomina el rastreo de referentes (74 %), seguido de lejos por la deixis temporal (11 %), deixis discursiva (8 %), y, por último, la presentación de participantes en el discurso (8 %). Recordemos que las instancias de deixis espacial fueron excluidas del análisis cuantitativo. Según el análisis de regresión, estas funciones no parecen tener un rol significativo en la concordancia o discordancia de género.

Aunque los tres grupos del estudio (bilingües, monolingües y hablantes de herencia) exhiben discordancia de género, el análisis de regresión solamente muestra como factor significativo a los monolingües, quienes favorecen la concordancia de género. Esto apoya la idea que la frecuencia de la discordancia de género tiene un vínculo con el contacto lingüístico entre el español y el kukama. Los monolingües, quienes son los únicos en no haber sido expuestos al kukama, muestran un patrón de concordancia más semejante al de variantes canónicas del español.

También es notable que, a pesar de que la extensión de las entrevistas es comparable para bilingües, monolingües y hablantes de herencia (38 036, 44 436 y 42 672 palabras, respectivamente), los bilingües produjeron la mitad de los demostrativos (50 %), mucho más que los otros grupos (25 % para monolingües y 25 % para hablantes de herencia). Debido a que los bilingües son los únicos que hablan el kukama con fluidez, esto podría ser un rasgo asociado a los bilingües que merece investigarse.

Como última observación, además del perfil sociolingüístico monolingüe, la única otra variable extralingüística que favorece la concordancia de género gramatical es el género social masculino de los participantes. Este resultado sigue patrones translingüísticos, en los que las mujeres lideran el cambio lingüístico en entornos tanto monolingües como bilingües cuando las innovaciones no son abiertamente estigmatizadas (ver Labov, 2001 y Lapidus Shin, 2013). En general, los fenómenos morfosintácticos son menos marcados (“below the level of consciousness”) en comparación con las variables léxicas y fonológicas (Silva-Corvalán & Enrique-Arias, 2017, p. 152). Los resultados del EAP también muestran que esta variedad sigue las tendencias documentadas para otras lenguas respecto a la función sintáctica de los demostrativos (Himmelmann, 1996, p. 214). Más específicamente, el uso adnominal es más frecuente (56 %) que el uso pronominal (44 %) en el EAP. Sin embargo, el tipo de función sintáctica del demostrativo es una de las variables significativas en predecir la concordancia de género. Como se vio en §6.4.2.2, la gran mayoría de los casos de discordancia (94 %) ocurren cuando el demostrativo tiene una función pronominal.

### 1.5.3 ¿Los patrones observados sugieren algún efecto de contacto?

Los resultados de los análisis de *Random Forest* y de regresión indican que el perfil sociolingüístico de los participantes es relevante en los patrones de concordancia. Este hallazgo sugiere un efecto de substrato lingüístico. Los datos tomados en su conjunto (Figura 2) revelan que las formas neutras en el corpus (42 %) son más frecuentes que las formas femeninas (25 %) y las formas masculinas (34 %). Además, las formas neutras en el EAP hacen referencia a proposiciones (52 %), tal como sucede en el español canónico, pero también a entidades, tanto masculinas (28 %) como femeninas (20 %). Esto significa que, en el EAP, el demostrativo *eso* parece estar conquistando el espacio funcional que le corresponde a las formas *ese* y *esa*. En otras palabras, *eso* podría estar evolucionando hacia un demostrativo general para funciones anafóricas, tanto para el rastreo de la referencia como para la presentación de participantes en el discurso.

Sin embargo, si observamos los datos más de cerca, los monolingües difieren significativamente de los no monolingües (bilingües y hablantes de herencia) en el uso innovador de *eso*, como lo revela el análisis de regresión. Específicamente, las instancias de discordancia (el uso de *eso* para referirse a entidades masculinas y femeninas) son menos frecuentes entre los hablantes monolingües de la ciudad (25 %) comparado con los hablantes de herencia (35 %) y los bilingües (40 %). Por otro lado, como muestra la Figura 7, la extensión funcional de *eso* ocurre significativamente con más frecuencia con entidades inanimadas (72 %), en comparación con entidades abstractas (13 %), animadas (8 %) y humanas (7 %). Nuestra lectura de estos resultados es que los hablantes de comunidades kukama, tanto bilingües como de herencia, muestran menor sensibilidad a la categorización de género de las entidades inanimadas.

Una posible explicación de esta extensión funcional innovadora de *eso* es que los hablantes de comunidades kukama categorizan entidades utilizando estrategias diferentes a las de los monolingües, y que estas estrategias emergen precisamente del sustrato lingüístico. Como se explicó en §6.2.1, en kukama se presta atención al género de quien habla, y en las prácticas discursivas tradicionales, todas las entidades animadas pueden participar en los “diálogos” a través de citas directas y siguiendo las reglas de género gramatical del idioma. Por ejemplo, si una narración incluye personajes como guacamayos, monos, y/o las “madres” de las plantas, cada uno participa en los diálogos según el género que le asigne el narrador, y este género se indexicaliza en citas directas a través de elementos gramaticales, incluyendo los demostrativos (para más detalles, ver Vallejos, 2015). Esta práctica discursiva orienta la atención de los hablantes a la animación de las entidades para su categorización en femeninas y masculinas, y su posterior rastreo a lo largo de la narrativa. Esta forma de construir el discurso kukama podría explicar por qué las entidades animadas en el corpus generalmente siguen patrones de concordancia canónicos, pero las entidades inanimadas son referidas en gran medida (72 %) a través del demostrativo neutro *eso*. Si nuestro análisis es correcto, esto significa que la distancia tipológica entre el kukama y el español desfavorece la concordancia entre el demostrativo y su anáfora cuando se trata de entidades inanimadas.

No obstante, los resultados tomados en su conjunto no apuntan al contacto lingüístico como el único responsable de la reorganización del sistema demostrativo en el EAP. Proponemos que los patrones de uso pueden ser explicados más bien como un proceso multicausal en el que el contacto lingüístico por substrato interactúa con la prominencia cognitiva de la animacidad para promover una reacomodación en el mapeo de los demostrativos y sus anáforas. Esta conclusión deberá ser evaluada a futuro a través de estudios de concordancia comparables, pero en comunidades con diferente sustrato.

#### 1.5.4 *Sistemas de género y el desplazamiento de lenguas*

Una reflexión final que queremos ofrecer se refiere a la estabilidad de los sistemas de género en los procesos de desplazamiento de lenguas en favor de lenguas mayoritarias como es, en este caso, el español. Las marcas de indexicalidad de género son las primeras candidatas a ser afectadas por la obsolescencia que está experimentando el kukama (Vallejos. 2016). Esto no es sorprendente dado que estudios recientes indican que los sistemas de indexicalidad de género son frágiles y propensos al cambio (Aikhenvald. 2016). De hecho, la indexicalidad de género parece ser mucho más sensible a los cambios socioculturales en comparación con el género gramatical. De ahí la relevancia de una documentación detallada de estas prácticas que pueda informar las iniciativas que buscan mantener una lengua viva.

Debido a que la transmisión natural del kukama ha sido interrumpida hace ya varias décadas, una pregunta crucial queda pendiente: ¿cómo adquirían los niños un sistema de género indexical? Algunos ancianos kukama cuentan que tanto hombres como mujeres aprendían primero la forma de hablar de la madre, pero que a la edad de 4 de 5 años los niños ya dominaban ambos generolectos. Estas observaciones concuerdan con lo que se sabe sobre otros sistemas, como el del karajá (Aikhenvald, 2016). Si bien por ahora no podemos responder con certeza la pregunta sobre adquisición, tal vez sea posible en el futuro. Al presente existen varias iniciativas para revitalizar el kukama que están siendo impulsadas por los mismos hablantes.

#### 1.6 *Conclusiones*

Los datos del corpus analizados en este estudio muestran que el sistema de demostrativos en el español amazónico de Perú (EAP) difiere significativamente de variantes canónicas del español. En primer lugar, no cuenta con demostrativos distales como *aquel* o *aquella*. La única distinción presente es entre próximos y medios, siendo estos últimos mucho más comunes que los primeros (con un ratio de 8:1). Adicionalmente, hay un notable grado de discordancia de género entre los demostrativos y sus anáforas para todos los tipos de hablantes, sean bilingües, hablantes de herencia o monolingües. Sin embargo, la discordancia en el EAP no ocurre de manera aleatoria. El nivel de discordancia está asociado con variables lingüísticas como el género del demostrativo, su rol sintáctico y el nivel de animacidad de la anáfora, además de variables no lingüísticas como el perfil sociolingüístico del hablante y su género social.

Uno de los hallazgos más importante es que el demostrativo medio neutro *eso* ha expandido sus funciones en el EAP; además de su uso canónico para referirse a proposiciones, puede emplearse para hablar de entidades de género masculino o femenino. Esta expansión es tal que, a diferencia de otras variantes del español, *eso* se ha convertido en el demostrativo más utilizado en el EAP. El uso innovador de *eso* varía entre los grupos de hablantes; su extensión es menor entre monolingües y mayor entre bilingües, mientras que los hablantes de herencia ocupan un lugar intermedio. La discordancia de género gramatical también está vinculada con la animacidad del referente, y ocurre principalmente cuando el referente del demostrativo es inanimado. La extensión funcional de *eso* posiblemente puede provenir de la interacción entre los sistemas radicalmente distintos de la expresión de género gramatical en el kukama y el español, la atención que los hablantes de kukama prestan a la animacidad de las entidades en el discurso tradicional, además de la prominencia cognitiva que tiene en general la animacidad. Esperamos que estos resultados motiven estudios sobre concordancia entre el demostrativo y su anáfora en otras variedades del español.

## Referencias bibliográficas

- Aikhenvald, A. Y. (2016). *How Gender Shapes the World*. Oxford: Oxford Academic Press.
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics* 24(1), 65-87.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>.
- Breiman, L. (2001). Random Forests. *Machine Learning*, (45), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Chafe, W. (1994). *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. (2007). Gender and noun classes. En T. Shopen (Ed.), *Language typology and syntactic description*, (Vol. 1): *Grammatical categories and the lexicon*, (2<sup>a</sup> ed.), (pp. 241-279). Cambridge: Cambridge University Press.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17(4), 463-489.
- Diessel, H., & Coventry, K. R.. (2020). Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review. *Frontiers in Psychology*, (11). DOI: 10.3389/fpsyg.2020.555265
- Fleming, L. (2012). Gender indexicality in the Native Americas: Contributions to the typology of social indexicality. *Language in Society*, 41(3), 295-320.

- Gundel, J. K., Hedberg, N., y Zacharski, R. (2004). Demonstrative pronouns in natural discourse. *Proceedings of the 5th Discourse Anaphora and Anaphor Resolution Colloquium* (Vol. 4), (pp. 81-86). São Miguel, Portugal.
- Haas, M. (1944). Men's and Women's Speech in Koasati. *Language*, 20(3), 142-149.
- Himmelmann, N. (1996). Demonstratives in narrative discourse: a taxonomy of universal uses. En B.A. Fox (Ed.), *Studies in Anaphora* (pp. 205-254). Amsterdam/Philadelphia: John Benjamins.
- Himmelmann, N. P., Sandler, M., Strunk, S., & Unterladstetter, V. (2018). On the universality of intonational phrases: a cross-linguistic interrater study. *Phonology*, 35(2), 207-245. doi:10.1017/S0952675718000039.
- Hinton, L. (2001). Language revitalization: An overview. En L. Hinton & K. L. Hale (Eds.), *The Green book of language revitalization in practice*, (pp. 3-18). Academic Press.
- Instituto Nacional de Estadística e Informática (INEI). (2017). *Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas*. <<http://censo2017.inei.gob.pe/>>.
- Labov, W. (2001). *Principles of Linguistic Change*, (Vol. 2): External factors. Oxford: Blackwell.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1): 45-79.
- Lapidus Shin, N. (2013). Women as Leaders of Language Change: A Qualification from the Bilingual Perspective. En Ana M. Carvalho & Sara Beaudrie (Eds.), *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, (pp. 135-147). Somerville, MA: Cascadilla Proceedings Project
- Lyons, J. (1977). Deixis, space and time. En *Semantics* (Vol. 2), (pp. 636-724). Cambridge: Cambridge University Press.
- Montrul, S. (2008). Review article: Second language acquisition welcomes the heritage language learner: opportunities of a new field. *Second Language Research*, SAGE Publications, 24(4), pp. 487-506.
- Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass*, 1(5), 368-395.
- Ramat, P. (2015). Language Change and Language Contact. En K. Jungbluth & F. D. Milano (Eds.), *Manual of Deixis in Romance Languages*, (pp. 581-596). De Gruyter Mouton.
- Rose, F. (2015). On male and female speech and more. A typology of categorical gender indexicality in indigenous South American languages. *International Journal of American Linguistics*, 81(4), 495-537.
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2), 155-163.

- Shin, N. & Vallejos-Yopán, R. (2023). Demostrativos y posesivos. En Guillermo Rojo, Victoria Vázquez Rozas and Rena Torres Cacoullos (Eds.), *Sintaxis del español*, (pp. 427-440). United Kindom: Taylor & Francis/Routledge.
- Silva-Corvalán, C., & Enrique-Arias, A. (2017). *Sociolingüística y pragmática del español* (2<sup>a</sup> ed.). Georgetown University Press.
- Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: opportunities lost or seized? *Modern Language Journal*, (89), 410-426.
- Vallejos, R. (2014). Peruvian Amazonian Spanish: Uncovering variation and deconstructing stereotypes. *Spanish in Context*, 11(3), 425-453
- Vallejos, R. (2015). La indexicalidad de género en kukama-kukamiria desde una perspectiva tipológica. En Ana Fernandez, Albert Alvarez & Zarina Estrada (Eds.), *Estudios de Lenguas Amerindias 3: contribuciones al estudio de las lenguas originarias de América*, (pp. 199-225). Hermosillo: Universidad de Sonora
- Vallejos, R. (2016). Structural outcomes of obsolescence and revitalization: documenting variation among the Kukama-Kukamirias. En Gabriela Perez-Baez, Chris Rogers y Jorge Rosés-Labrada (Eds.), *Language Documentation and Revitalization in Latin America*, (pp. 143-164). Berlin: De Gruyter Mouton.
- Vallejos-Yopán, R. (2023). From demonstrative to filler: Este in Amazonian Spanish and beyond. *Linguistics*, 61(3), 651-678. <https://doi.org/10.1515/ling-2021-0137>.
- Vallejos-Yopán, R., García-Rivera, F., & Rosales-Alvarado, H. (2022). Indigenous languages in higher education: case studies from the Amazon of Peru. En Rosita L. Rivera & Eva Rodríguez-González (Eds.), *Integrating Context-based Approaches to Language Assessment in Multilingual Settings*, (pp. 53-191). Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.6762282>
- Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Tatiana Nikitina, M. Catherine O'Connor, & Tom Wasow. (2004). Animacy encoding in English: why and how. En *Proceedings of the Association for Computational Linguistics Workshop on Discourse Annotation*, (pp. 118-125).