

Marcación morfosintáctica de las categorías de tiempo, aspecto y modo (TAM) en lenguas de Sudamérica: Una aproximación preliminar.

Tatiana Maldonado Muñoz
th.mmunoz@gmail.com
Universidad de Chile

1. INTRODUCCIÓN

Sudamérica tiene una gran diversidad de lenguas indígenas, las cuales no están necesariamente relacionadas entre sí y entre las que pueden establecerse diferencias a nivel estructural (Adelaar & Muysken, 2004). Sin embargo, también existen lenguas con rasgos lingüísticos compartidos (sean de una misma familia lingüística o de diferentes), lo que podría ser producto del contacto entre lenguas. Determinar dichos rasgos puede contribuir a definir áreas lingüísticas, es decir, zonas donde se concentran lenguas con características comunes (Campbell, 2007). En este trabajo, el interés se centra en las lenguas de Sudamérica que cuentan con marcadores morfosintácticos para las categorías de tiempo, aspecto y modo. Nuestra investigación busca responder las preguntas: a) ¿cuáles son las lenguas de Sudamérica que marcan morfosintácticamente las categorías de tiempo, modo y aspecto? y b): Considerando las lenguas con marcadores morfosintácticos para TAM ¿cómo pueden agruparse las lenguas geográficamente y cuál es la distancia tipológica entre estas y lenguas del mismo territorio sin marcas morfosintácticas?

Para responder estas interrogantes, nos proponemos clasificar las lenguas indígenas de Sudamérica de acuerdo a la presencia o ausencia de marcadores morfosintácticos para codificar las categorías de tiempo, aspecto y modo (TAM) a través de técnicas computacionales. Posteriormente, en base a los resultados obtenidos, se buscará definir grupos de lenguas por área geográfica estableciendo la distancia tipológica entre lenguas de una misma área mediante un análisis computacional.

Como hipótesis inicial proponemos que las lenguas de Sudamérica tienden a no marcar morfosintácticamente las categorías TAM, siendo una característica general de esta zona.

Como segunda hipótesis planteamos que puede establecerse la distancia tipológica entre lenguas que tienen marcadores morfosintácticos y aquellas que no, la cual podría observarse dentro de una misma área geográfica o familia lingüística.

Los resultados obtenidos son una aproximación preliminar para agrupar lenguas con rasgos compartidos para las categorías de TAM en Sudamérica y busca aportar a investigaciones futuras para definir posibles áreas lingüísticas. Por otra parte, las estrategias computacionales empleadas incorporan nuevas metodologías para el análisis de los datos lingüísticos, ampliando el alcance del trabajo realizado

2. MARCO CONCEPTUAL

2.1. TAM: tiempo, aspecto y modo

La sigla TAM es la abreviación para las categorías de tiempo, aspecto y modo (Dahl, 2015) y serán abarcadas desde su dimensión gramatical.

El tiempo es una expresión gramaticalizada de la información deíctica sobre la localización temporal de un determinado evento (Klein, 2009; RAE y ASALE, 2010). Si bien las lenguas pueden expresar esta categoría a través de diversas formas, nos delimitaremos al tiempo gramatical y sus distinciones de: presente, pasado y futuro. Al interactuar el punto de habla (momento en que la oración es producida) y el punto del evento (momento en que la acción referida en la oración sucede) se da una relación temporal. Al tomar el punto de habla como referencia, tenemos que: a) si el evento ocurre antes, corresponde a pasado, si ocurre simultáneamente, corresponde a presente y si ocurre posteriormente, corresponde a futuro (Muller, 2013; Reichenbach, 1984).

Por su parte, la categoría de aspecto -en oposición a la de tiempo- corresponde a la

información temporal no deíctica sobre cómo se desarrolla y distribuye un evento en el tiempo (Bosque & Demonte, 1999; García, 1998). Se suele distinguir entre aspecto léxico y aspecto gramatical, pues el primero se relaciona con la información aspectual inherente que aporta el verbo, mientras que el segundo, se refiere a la manifestación morfológica de las diversas formas en que se puede observar un evento (Comrie, 1989). Dentro del aspecto gramatical, se realiza una clasificación general entre: aspecto perfectivo (si el evento puede visualizarse de principio a fin) y aspecto imperfectivo (si solo puede visualizarse una parte interna del evento) (Smith, 1997).

Por último, el modo es una categoría gramatical relacionada con el estado de la proposición que describe el evento. Una distinción es la de *realis* e *irrealis*, aunque controversial dentro del dominio de la modalidad, permite obtener información de lo que un hablante percibe como real o irreal. De esta forma, el modo *realis* retrata situaciones actualizadas, ocurridas o que están ocurriendo, las cuales son conocidas a través de la percepción directa. El modo *irrealis*, retrata situaciones que son propias del pensamiento, conocidas a través de la imaginación (Palmer, 2001).

2.2. Marcadores

Cuando una categoría tiene una marca con contenido morfológico para expresar dicha categoría, se emplea el concepto de marcador. Por tanto, para efectos de los datos analizados un marcador podrá ser un afijo, clítico, partícula o un auxiliar que posea información de tiempo, aspecto o modo (Muller, 2013).

3. MARCO METODOLÓGICO

3.1. Distancia tipológica

La distancia tipológica es un concepto relevante para la última etapa del análisis de los datos de este trabajo. Corresponde a la diferencia tipológica entre dos o más lenguas (de una misma familia lingüística o diferente) y permite medir cuantitativamente la distancia existente

entre estas lenguas respecto a un mismo rasgo (Chiswick & Miller, 2004). Como la mayoría de las bases de datos se conforma de un conjunto de lenguas y rasgos pueden establecerse valores para ser asignados a dichos rasgos y lenguas. Es posible expresar esto a través de la fórmula $L_X[i]$, donde X es la lengua y $[i]$ es el valor del rasgo a analizar. Lo anterior permite ver qué tan similares o disímiles son las lenguas en un rango de 0.0 (idénticas) a 1 (completamente distintas) y puede calcularse empleando la fórmula llamada distancia de Hamming (ver Figura 1) (Hammarström & O'Connor, 2013).

$$G(L_X, L_Y) = \frac{\#_{i \in DEF(L_X, L_Y)} L_X[i] \neq L_Y[i]}{|DEF(L_X, L_Y)|}$$

Figura 1. Fórmula distancia de Hamming.

4. METODOLOGÍA

4.1. Los datos y la muestra

Nuestro trabajo se basa en la sistematización de rasgos TAME proporcionados por Müller (2016), datos disponibles en The South American Indigenous Language Structures (SAILS). Se empleó la muestra definida por la autora conformada por 64 lenguas indígenas de Sudamérica (ver anexo 1). Las familias lingüísticas de la muestra y el número de lenguas por familia, se presentan en la Tabla 1 (la distribución geográfica de las lenguas se presenta en el mapa del Anexo 2).

FAMILIAS LINGÜÍSTICAS		
Arahuaca (5)	Chocó (1)	Paez (1)
Arahuana (1)	Chono (2)	Panoa (3)
Araucana (1)	Guaycurú (2)	Quechua (3)
Aymara (1)	Jívaras (1)	Tacana (1)
Barbacoana (2)	Macro-Ye (4)	Tucana (2)
Borán (2)	Nadahup (2)	Tupí (9)
Caribe (3)	Matacoan (1)	Yanomana (1)
Chapacura (1)	Mosetena (1)	Aisladas (11)
Chibcha (2)	Nambikwára (2)	

Tabla 1. Familias lingüísticas de la muestra.

4.1.1. Rasgos lingüísticos de la muestra

Para la sistematización realizada por Müller (2013) para las categorías TAME, define una

serie de rasgos lingüísticos, los que están disponibles en SAILS. Para cada uno de los rasgos, la autora definió preguntas en relación a si presentan marcadores morfosintácticos (afijos, clíticos, partículas, auxiliares o reduplicación en algunos casos). Para este trabajo, los datos que se consideraron para el análisis son de las primeras tres categorías (TAM) y se seleccionaron los rasgos especificados en la Tabla 2.

RASGOS LINGÜÍSTICOS		VALORES
1. Tiempo	1.1 Presente 1.2 Pasado 1.3 Futuro	Marcado No marcado
2. Aspecto	2.1 Perfectivo 2.2 Imperfectivo	
3. Modo	3.1 Realis 3.2 Irrealis	

Tabla 2. Rasgos lingüísticos de la muestra.

4.2. Técnicas computacionales para el procesamiento y análisis de los datos

Para la conformación y análisis del *dataframe* del presente trabajo, se realizaron los siguientes pasos:

- 1) se descargaron los datos asociados a nuestra muestra y los rasgos lingüísticos definidos en la sección anterior. Cada rasgo y sus respectivos datos, se registraron en archivos individuales para facilitar su análisis en formato .csv¹.
- 2) se empleó la aplicación JupyterLab para visualizar cada uno de los archivos, con ayuda de la librería Pandas. Con la ayuda de la misma aplicación, se llevaron a cabo los pasos que se describen a continuación.
- 3) se realizó una separación de rasgos lingüísticos por categoría. De esta forma, los datos se organizaron en: tiempo (pasado, presente y futuro), aspecto (perfectivo e imperfectivo) y modo (realis e irrealis).
- 4) para identificar y agrupar las lenguas conforme a los criterios de marcación/no

marcación, se crearon diccionarios para almacenar la información por cada rasgo lingüístico. En un diccionario se guardaron las lenguas con marcadores morfosintácticos, y en otro, aquellas que carecían de estos.

- 5) se elaboró una tabla para cada categoría, en la cual se visualizan el número total de lenguas con marcación y no marcación del rasgo lingüístico para facilitar el análisis de los datos.
- 6) Para cada rasgo se creó un mapa para graficar la distribución de las lenguas conforme a las características marcado/no marcado.

5. Resultados preliminares

Los resultados obtenidos permiten comprobar nuestra primera hipótesis. Efectivamente, existe una tendencia a la no marcación de las categorías TAM en la mayoría de los rasgos analizados, pero también existen rasgos con una tendencia opuesta.

En la Tabla 3 se presentan los resultados obtenidos para la categoría de tiempo y sus tres rasgos analizados: presente, pasado y futuro. Las lenguas se agruparon conforme al rasgo marcado y no marcado. En la sección 7, se incluyen en forma de anexo los mapas para graficar la distribución de las lenguas por cada rasgo lingüístico. Las lenguas marcadas con amarillo son las que presentan algún tipo de marcador morfosintáctico y las marcadas con rojo, son las que no presentan marcación para el rasgo.

TIEMPO			
RASGO	Presente	Pasado	Futuro
Marcado	18 (28%)	45 (70%)	55 (86%)
No marcado	46 (72%)	19 (30%)	9 (14 %)

Tabla 3. Distribución de lenguas para los rasgos de la categoría Tiempo.

Respecto al rasgo presente (ver Anexo 3) vemos que existe un 28% (18 lenguas) que marca

¹ Los archivos y códigos empleados para el análisis de los datos, se encuentran disponibles en el siguiente link:

morfosintácticamente presente y un 72% (46 lenguas) que no lo marca. En el siguiente rasgo correspondiente a pasado (ver Anexo 4), se observa una tendencia diferente: el 70% (45 lenguas) marca morfosintácticamente el rasgo y un 30% (19 lenguas) no lo marca. Para el rasgo futuro (ver Anexo 5) ocurre algo similar, un 86% (55 lenguas) marca morfosintácticamente el rasgo y el 9% (14 lenguas) no lo marca. Esta categoría presenta la particularidad de presentar dos rasgos (pasado y futuro) que se alejan de nuestra hipótesis y será relevante tomar estos resultados para analizarlos posteriormente. Por el momento, podemos señalar que: futuro y pasado es el rasgo más codificado en las lenguas de Sudamérica, mientras que, el presente es el con menor grado de codificación.

En la categoría de aspecto, los resultados obtenidos muestran una distribución equitativa de las lenguas en los rasgos, es decir, existe un mismo número de lenguas que marca perfecto e imperfecto como se observa en la Tabla 4 (ver anexo 6 y 7).

ASPECTO		
RASGO	Perfectivo	Imperfectivo
Marcado	25 (39%)	25 (39%)
No marcado	39 (61%)	39 (61%)

Tabla 4. Distribución de lenguas según el rasgo marcado/no marcado para la categoría Aspecto.

Los resultados de la tabla anterior, muestran una mayor tendencia a la no marcación de aspecto, tanto para el rasgo perfecto como imperfecto. Un 39% (25 lenguas) presenta algún tipo de marcador morfosintáctico para aspecto perfecto e imperfecto y un 61% (39 lenguas) no marca estos rasgos.

Por último en la Tabla 5, se pueden observar los resultados obtenidos para la categoría de modo.

MODO		
RASGO	Realis	Irrealis
Marcado	8 (12.5%)	30 (47%)
No marcado	56 (87.5%)	34 (53%)

Tabla 5. Distribución de lenguas según el rasgo marcado/no marcado para la categoría Modo.

En el primer rasgo de la categoría correspondiente a realis (ver Anexo 8Anexo 8), existe un 12,5% (8 lenguas) con marcación morfosintáctica del rasgo y un 87,5% (56 lenguas) que no lo marcan. Para el rasgo irrealis (ver Anexo 9), hay un 47% (30 lenguas) con marcación morfosintáctica y un 53% (34 lenguas) sin marcación para este rasgo. Cabe destacar que, el rasgo realis es el que presenta la diferencia más alta en codificación respecto a los rasgos de todas las categorías. Así, se constituye como el rasgo menos codificado en las lenguas de Sudamérica analizadas.

En síntesis, los rasgos lingüísticos analizados quedan organizados de la manera en que se muestra en la Tabla 6, conforme al menor porcentaje de marcación al mayor.

RASGO	PORCENTAJE DE LENGUAS CON MARCADORES MORFOSINTÁCTICO
Modo Realis	12.5%
Tiempo Presente	28%
Aspecto Perfectivo e Imperfectivo	39%
Modo Irrealis	47%
Tiempo Pasado	70%
Tiempo Futuro	86%

Tabla 6. Rasgos lingüísticos ordenados por porcentaje de marcación morfosintáctica.

En base a estos resultados, nos parece pertinente centrarnos en los rasgos lingüísticos donde se observa una mayor diferencia en la codificación morfosintáctica: pasado y futuro. Tras realizar estos análisis descriptivos, procederemos a realizar otros para responder nuestra segunda interrogante. Por tanto, proyectamos los siguiente: a) agrupar las lenguas por zonas geográficas conforme a la marcación/no marcación de los rasgos en general y una vez definidos los grupos. Para este propósito, realizaremos una visualización en dos dimensiones de los grupos de lenguas con ayuda de la librería Seaborn u otra similar. b) se calculará la distancia tipológica entre las lenguas que codifican los rasgos seleccionados

(pasado y futuro) de una determinada zona con aquellas lenguas de la misma zona sin marcadores morfosintácticos para estos rasgos, empleando un código para calcular distancia de Hamming.

Para cerrar este apartado, es necesario señalar que al realizar los códigos que permitieron la elaboración de estos análisis y de los mapas anexados que los acompañan, debe existir una rigurosidad en la escritura de los códigos, pues se trabaja con varios rasgos para una misma categoría. Equivocarse en el nombre de un diccionario, por ejemplo, causará un error o datos no correspondientes al rasgo.

6. Conclusiones y perspectivas futuras

A partir de los primeros análisis podemos concluir lo siguiente:

En las categorías de aspecto y modo se cumple nuestra hipótesis inicial sobre la tendencia a la no marcación de estas categorías en las lenguas de Sudamérica. En algunos rasgos se observa una mayor diferencia que en otras.

Respecto a la categoría de tiempo, solo el rasgo presente se adhiere a la hipótesis inicial. Los rasgos de pasado y futuro muestran una tendencia a la marcación morfosintáctica, posiblemente se encuentran más gramaticalizados en las lenguas. La explicación de este fenómeno debe ser investigado en trabajos posteriores.

En el análisis posterior es necesario considerar la distribución geográfica de las lenguas para los rasgos para determinar zonas geográficas donde se concentran las lenguas con mayor o menor marcación de un rasgo, pues puede ser una posible explicación para el fenómeno reportar.

Las técnicas computacionales empleadas han permitido alcanzar los objetivos propuestos hasta ahora. Sin embargo, deben elaborarse los códigos cuidadosamente para evitar errores, especialmente al trabajar con varios rasgos en una misma categoría.

Por último, debe señalarse que, este trabajo es una primera aproximación para agrupar las lenguas consideradas en nuestro análisis conforme a la marcación de TAM, lo que en investigaciones futuras podría ayudar a definir áreas lingüísticas.

7. Bibliografía

- Adelaar, W., & Muysken, P. (2004). *The languages of the Andes*. Cambridge University Press.
- Bosque, I., & Demonte, V. (1999). *Gramática descriptiva de la lengua española 2. Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. (Vol. II). Espasa.
- Campbell, L. (2017). Why is it so Hard to Define a Linguistic Area? In R. Hickey (Ed.), *Handbook of Areal Linguistics* (pp. 19-39). Cambridge University Press.
- Chiswick, B., & Miller, P. (2004). *Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages*. IZA.
- Dahl, Ö. (2015). Tense, Aspect, Mood and Evidentiality. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 210-213). Elsevier.
- García, L. (1998). *El aspecto gramatical en la conjugación*. Arco/Libros.
- Hammarström, H., & O'Connor, L. (2013). Dependency-sensitive typological distance. In S. Anju & B. Lars (Eds.), *Approaches to Measuring Linguistics Differences*. De Gruyter.
- Klein, W. (2009). How time is encoded. In W. Klein & L. Ping (Eds.), *The Expression of Time* (pp. 39-82). De Gruyter Mouton.
- Muller, N. (2013). *Tense, Aspect, Modality and Evidentiality Marking in South American Indigenous Languages*. LOT.
- Müller, N. (2016). Tense-Aspect-Mood-Evidentiality (TAME). In P. e. a. e. Muysken (Ed.), *South American Indian Language Structures (SAILS) Online*. Max Planck Institute for the Science of Human History. <https://sails.cild.org>
- Palmer, F. (2001). *Mood and Modality*. Cambridge University Press.
- RAE y ASALE. (2010). *Nueva gramática de la lengua española. Manual*. Espasa.
- Reichenbach, H. (1984). *Elements of Symbolic Logic*. The MacMillan Company.
- Smith, C. (1997). *Parameter of aspect*. Kluwer Academic Publisher.

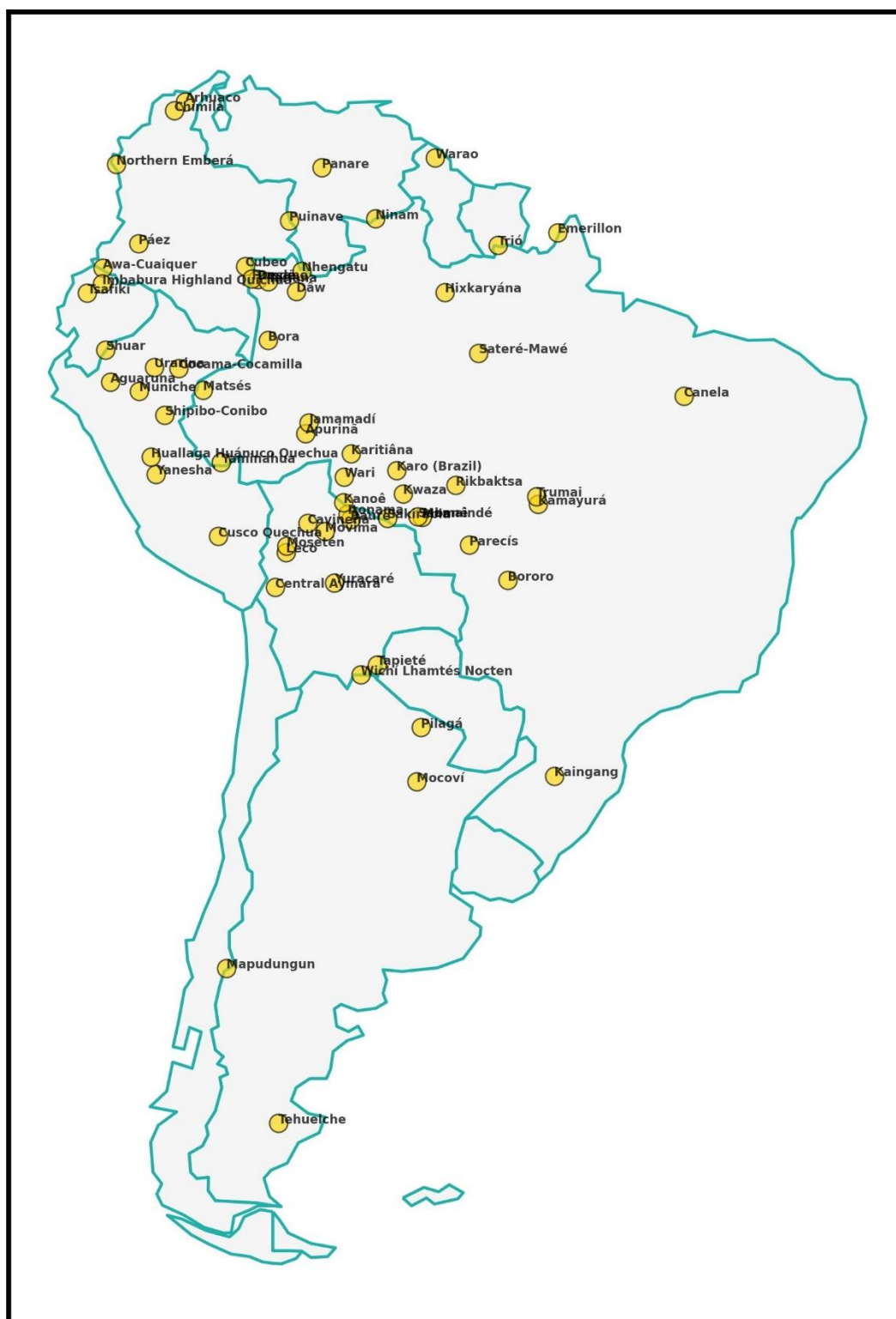
8. ANEXOS

FAMILIA	LENGUA
Arahuaca	Apurinã Baure Tariana Yanesha Paresí
Arahuana	Taíno (Jarawara)
Araucana	Mapudungun
Aymara	Aymara
Barbacoanas	Awá pit Tsafiki
Borán	Miraña
Caribe	Hixkaryána Panare Tiriyó
Chapacura	Wari
Chibchas	Chimila Iku
Chocó	Emberá
Chono	Tehuelche
Guaicurú	Mocoví Píлага
Jívaras	Aguaruna
Macro-ye	Bororo Rikbaktsá Káingang Timbira
Nadahup	Dâw Hup
Matacoan	Wichí
Mosetenas	Mosetén
Nambikwára	Mamaindê Sabanê
Paez	Nasa Yuwe
Pano	Matsés Shipibo Yaminahua
Quechua	Huallaga Quechua Imbabura Quechua Cuzco Quechua
Tacana	Cavineña
Tucana	Cubeo Desano
Tupí	Karitiana Karo Mekéns Sateré-Mawé Emérillon Kamaiurá Cocama-Cocamilla Tapiete Nheegatú
Yanomana	Yanam

Aisladas/Sin clasificar	Itomana Leko Movima Yurakaré Kanoê Kwaza Trumai Munichi Puinave Urarina Warao
-------------------------	---

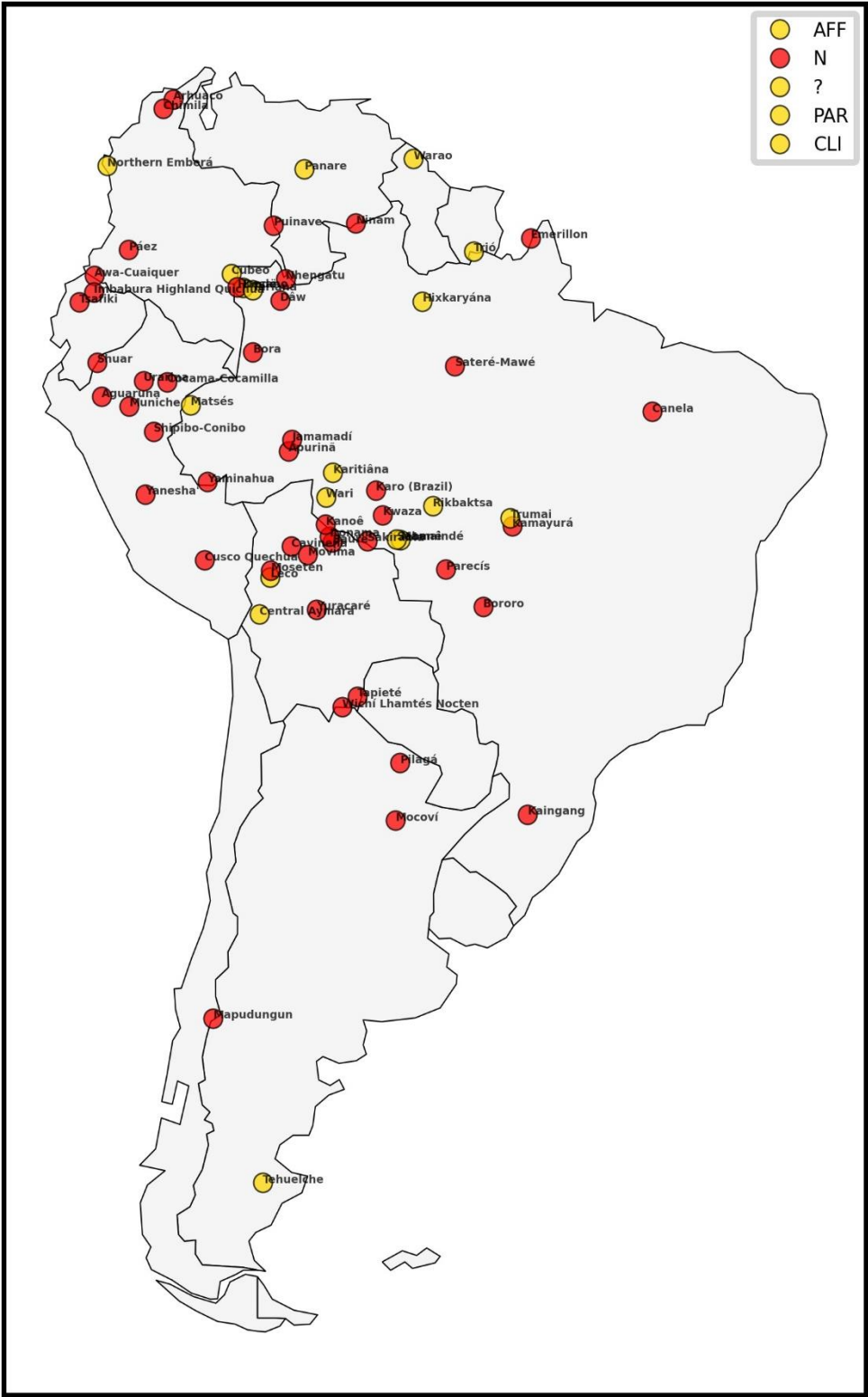
Anexo 1. Muestra: Lenguas por familias lingüísticas

Lenguas analizadas



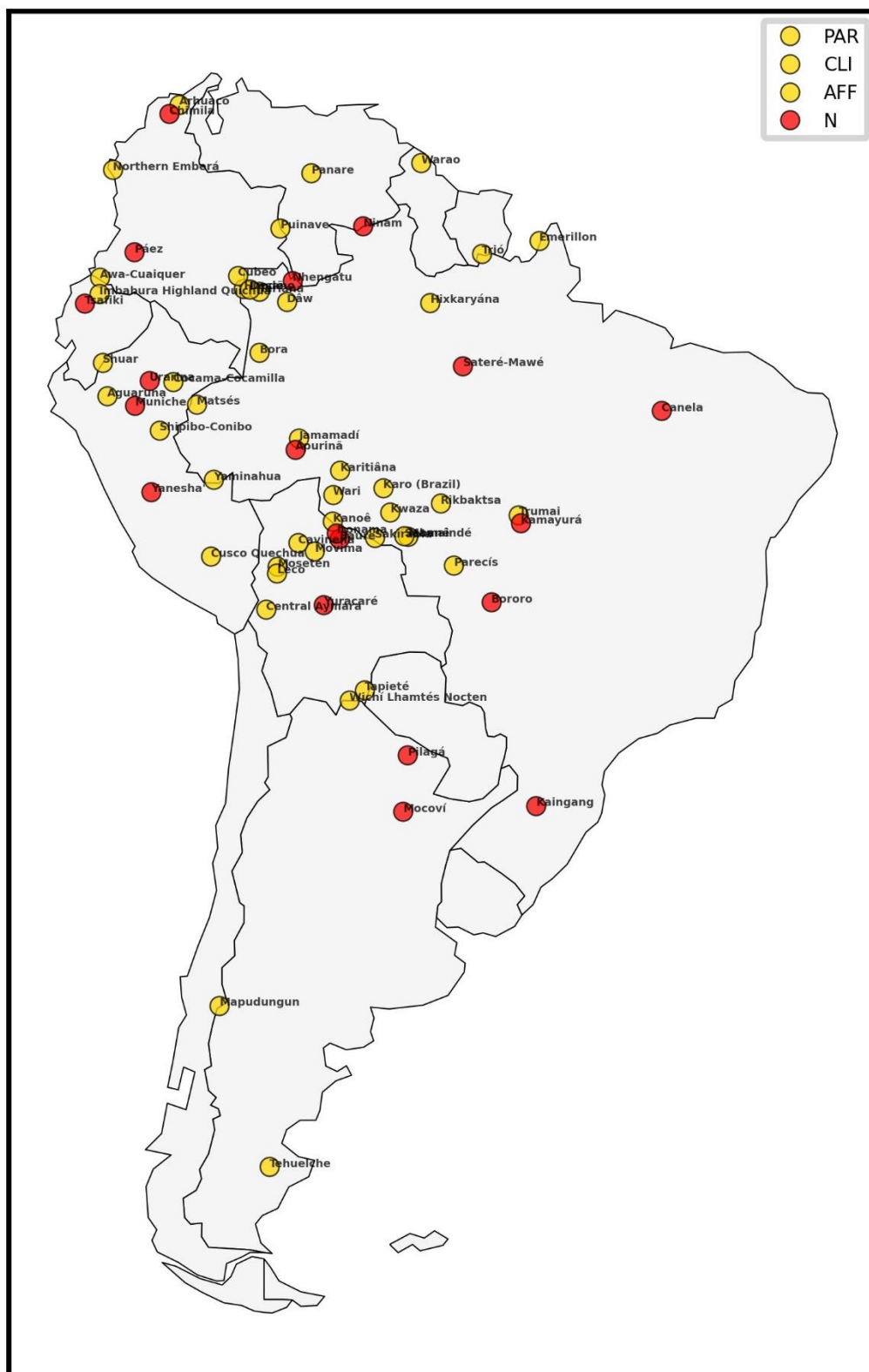
Anexo 2. Mapa Lenguas Analizadas. Distribución geográfica de lenguas de la muestra.

Marcación Presente



Anexo 3. Mapa Marcación Presente. Distribución de lenguas con marcación morfosintáctica para tiempo presente.

Marcación Pasado



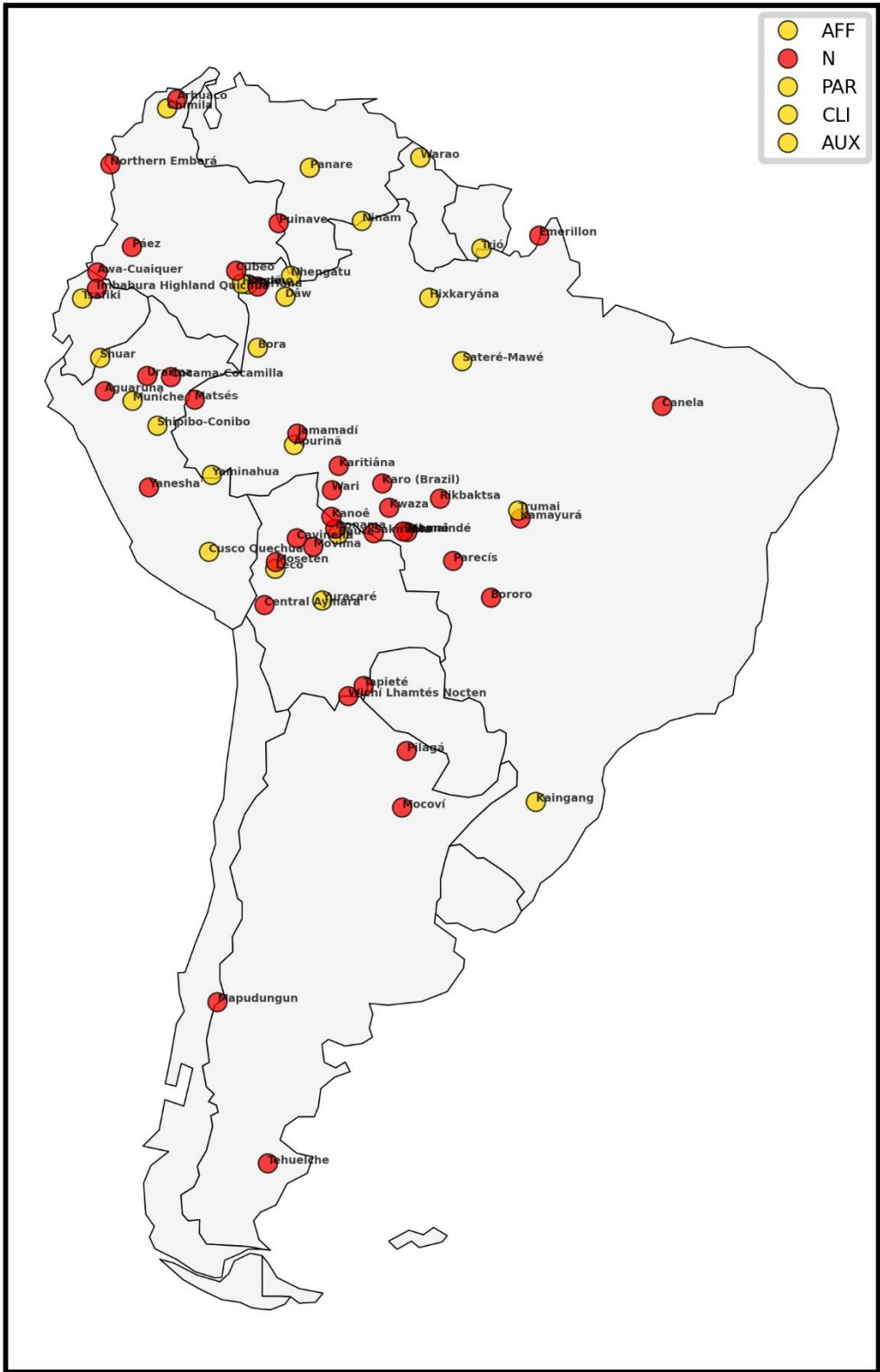
Anexo 4. Mapa Marcación Pasado. Distribución lenguas con marcas morfosintácticas para tiempo pasado.

Map of South America showing the distribution of 40 languages. The map is color-coded by language family: AFF (yellow), N (red), PAR (orange), CLI (light green), and AUX (dark green). The languages are distributed across the continent, with a high concentration in the northern and central regions. The legend in the top right corner identifies the color coding for the language families.

Language Family	Language
AFF	Achuar, Comilla, Northern Emborá, Panare, Warao, Emerillon, Unin, Hixkaryana, Sateré-Mawé, Canela, Karitiána, Wari, Karo (Brazil), Rikbaktsa, Kanoé, Kwaza, Kikandé, Parecís, Bororo, Yuki, Lhamtés, Nocten, Kaingang, Tehuelche
N	Pse, Awa-Cuaiquer, Inbikura Highland, Shuar, Yanoama, Cocamilla, Aguaruna, Muniche, Matsés, Shilbo-Conibo, Yanesha, Yaminahua, Karitiána, Wari, Karo (Brazil), Rikbaktsa, Kanoé, Kwaza, Kikandé, Parecís, Bororo, Yuki, Lhamtés, Nocten, Kaingang, Tehuelche
PAR	Unin, Hixkaryana, Sateré-Mawé, Canela, Karitiána, Wari, Karo (Brazil), Rikbaktsa, Kanoé, Kwaza, Kikandé, Parecís, Bororo, Yuki, Lhamtés, Nocten, Kaingang, Tehuelche
CLI	Shuar, Yanoama, Cocamilla, Aguaruna, Muniche, Matsés, Shilbo-Conibo, Yanesha, Yaminahua, Karitiána, Wari, Karo (Brazil), Rikbaktsa, Kanoé, Kwaza, Kikandé, Parecís, Bororo, Yuki, Lhamtés, Nocten, Kaingang, Tehuelche
AUX	Mapudungun

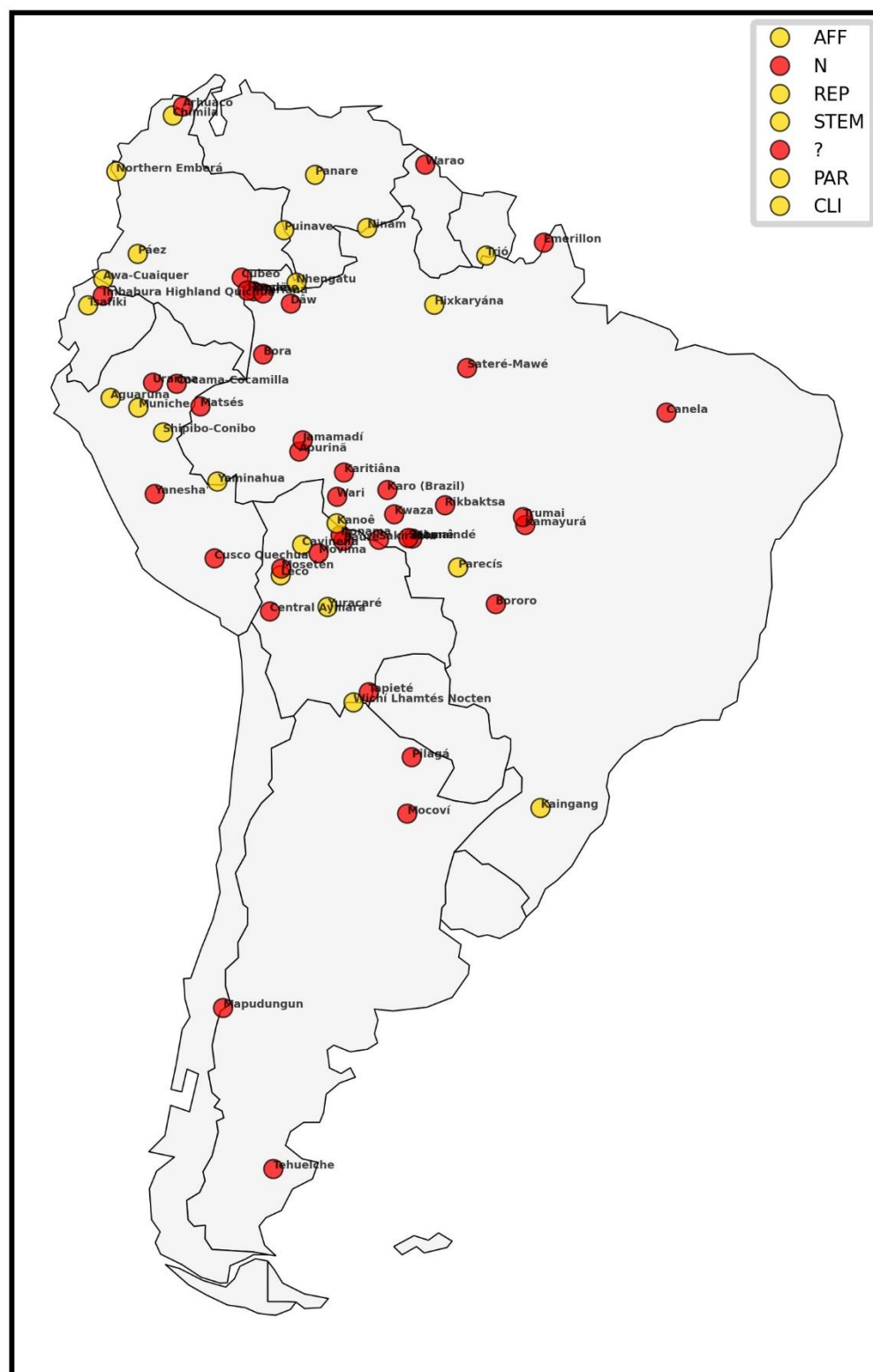
Anexo 5. Mapa Marcación Futuro. Distribución lenguas con marcas morfosintácticas para tiempo futuro.

Marcación Perfectivo



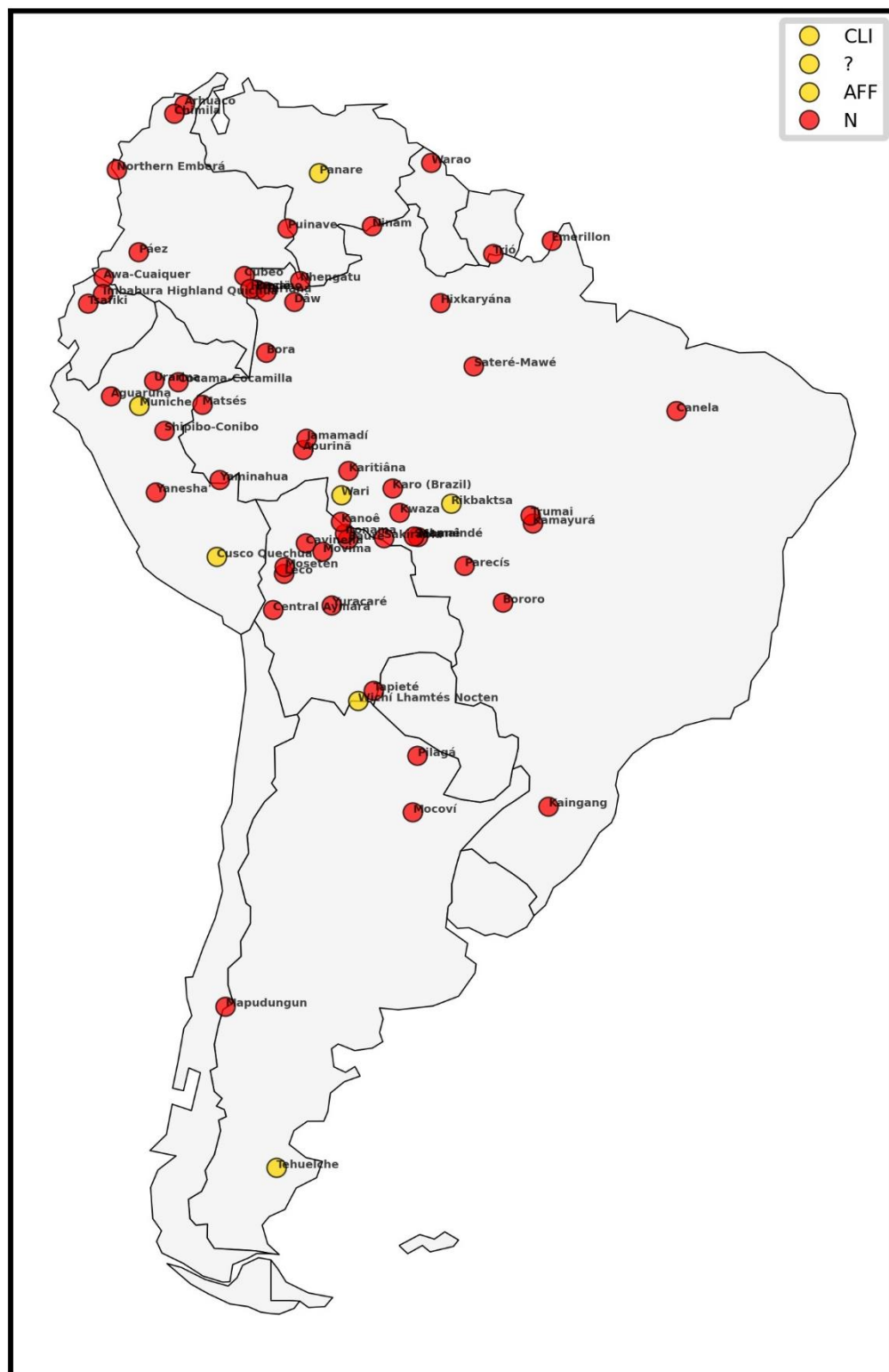
Anexo 6. Mapa Marcación Perfectivo. Distribución lenguas con marcas morfosintácticas para aspecto perfectivo.

Marcación Imperfectivo



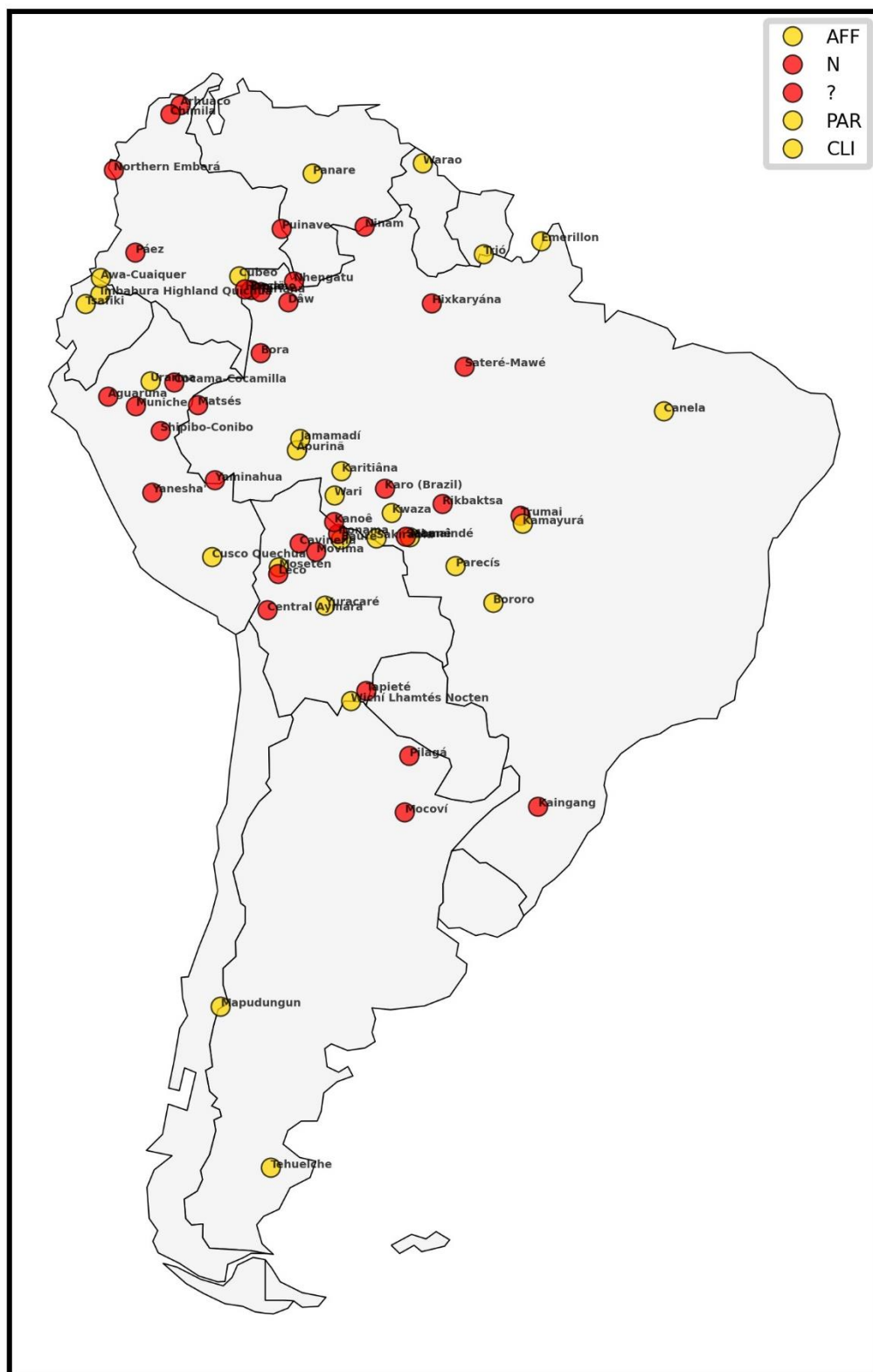
Anexo 7. Mapa Marcación Imperfectivo. Distribución lenguas con marcas morfosintácticas para aspecto imperfectivo.

Marcación Realis



Anexo 8. Mapa Marcación Realis. Distribución lenguas con marcas morfosintácticas para modo realis.

Marcación Irrealis



Anexo 9. Mapa Marcación Irrealis. Distribución lenguas con marcas morfosintácticas para modo irrealis.