

# Introducción a la Tipología (y la Tipología Computacional)

June 6, 2025

- 1 Fundamentos de la Tipología Lingüística
  - Definiendo la Tipología Lingüística
- 2 Tipología vs. Lingüística Histórica vs. Areal
- 3 Conceptos Centrales en Tipología Lingüística
  - Universales Lingüísticos
  - Comparación Interlingüística y Espacio Tipológico
- 4 Ejemplo de Subcampo y Cuestiones Prácticas
  - Tipología Sintáctica
  - Muestreo de Lenguas
  - Bases de Datos Tipológicas

## Definiendo la Tipología Lingüística

# Qué es la Tipología Lingüística?

- **Definición:** La tipología lingüística (o tipología de lenguas) es un campo de la lingüística que estudia y clasifica las lenguas según sus rasgos estructurales para permitir su comparación.
- **Objetivo Principal:** Describir y explicar la diversidad estructural observada en las lenguas del mundo e identificar sus propiedades comunes, a menudo denominadas universales lingüísticos.

## Metas:

- Identificar los posibles tipos de estructuras lingüísticas que se encuentran en las lenguas humanas.
- Comprender el rango completo y los límites inherentes de la variación lingüística.
- Descubrir patrones lingüísticos universales—que son comunes a muchas, o incluso a todas, las lenguas.
- Explicar por qué ciertos patrones estructurales son comunes, por qué algunos son raros y por qué otros podrían no estar atestiguados.

## Tipología de Todo:

- Abarca todos los niveles de análisis lingüístico:
  - Fonética y Fonología (sistemas de sonidos y su organización).
  - Morfología (estructura interna de las palabras).
  - Sintaxis (estructura de la oración, orden de palabras y relaciones gramaticales).
  - Semántica (significado, incluyendo el significado léxico y gramatical).
  - Léxico (patrones y organización del vocabulario).
- Este amplio alcance también permite la interacción, donde los patrones en un nivel (p. ej., el marcaje de caso morfológico) pueden interactuar con o predecir patrones en otro (p. ej., el orden de palabras sintáctico).

# Rol Amplio de la Tipología Lingüística

El estudio de la tipología lingüística es importante en todo el campo de la lingüística:

- **Cognición del Lenguaje Humano:** Revelar ideas sobre la naturaleza de la facultad humana del lenguaje, incluyendo posibles restricciones cognitivas y presiones funcionales que moldean la estructura del lenguaje.
- **Teoría Lingüística:** Pone a prueba las teorías lingüísticas en la medida en que hacen predicciones sobre qué tipos de gramáticas pueden tener las lenguas y cómo interactúan los módulos de esas gramáticas.
- **Documentación y Revitalización de Lenguas:** Predice rasgos en lenguas poco documentadas o en peligro de extinción, también informando documentación y proyectos de revitalización.
- **Lingüística Histórica:** Aunque es distinta de la clasificación genealógica, los patrones tipológicos pueden ofrecer pistas sobre las vías del cambio lingüístico y las posibles conexiones históricas entre las lenguas.

- Joseph Greenberg (1915-2001) es considerado como el fundador de la tipología lingüística moderna (aunque hubo tipólogos en el siglo XIX y XVIII (p. ej., Gabelentz, Jakobson)).
  - Es imprescindible conocer: *Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements* (1963).
  - Formalizó el concepto de **universales implicacionales** (p. ej., "Si una lengua tiene un orden dominante VSO, casi siempre tiene preposiciones").
  - Enfatizó el uso de muestras de lenguas grandes y diversas y métodos cuantitativos para identificar tendencias estadísticas.



# Aislando la Tipología de otras Fuentes de Similitud

# Tipología vs. Lingüística Histórica vs. Areal

Es crucial distinguir la tipología lingüística de la lingüística histórica y areal:

- Tipología Lingüística: Queremos descubrir aquellos aspectos del lenguaje que son universales porque **así son las lenguas humanas**.
- Lingüística Histórica:
  - Clasifica las lenguas basándose en el **parentesco histórico** y la ascendencia común, agrupándolas en familias lingüísticas (p. ej., indoeuropeo, maya, uto-azteca).
  - Claramente, las lenguas pueden tener propiedades similares por la razón que tienen una lengua ancestral compartida.

La investigación tipológica requiere un muestreo cuidadoso para evitar la **autocorrelación**, asegurando que los patrones observados no sean solo artefactos de unas pocas familias lingüísticas grandes. Esto necesita el uso de información de la lingüística histórica para construir muestras equilibradas.

# Tipología vs. Lingüística Histórica vs. Areal

Es crucial distinguir la tipología lingüística de la lingüística histórica y areal:

- Tipología Lingüística: Queremos descubrir aquellos aspectos del lenguaje que son universales porque **así son las lenguas humanas**.
- Lingüística Areal:
  - Clasifica las lenguas basándose en el **contacto**, agrupándolas en áreas lingüísticas (p. ej., Área Lingüística Mesoamericana).
  - Claramente, las lenguas pueden tener propiedades similares en virtud del hecho de que están en contacto, y por lo tanto se vuelven similares entre sí.

La investigación tipológica requiere un muestreo cuidadoso para evitar el efecto de **las áreas lingüísticas**, es decir, que las lenguas sean similares en tipo debido al contacto, no porque las lenguas tiendan a tener ese tipo. Esto necesita el uso de información sobre el contacto lingüístico para construir muestras equilibradas.

# Tipología vs. Lingüística Histórica vs. Areal

Estos son campos distintos pero complementarios.

- En particular, los desarrollos históricos y las vías de cambio lingüístico, estudiados por la lingüística histórica, a menudo conducen a las distribuciones tipológicas observadas hoy en día.
- De manera similar, esperamos que las lenguas, cuando están en contacto, tomen prestado unas de otras de maneras que sean tipológicamente válidas (o si no, esperamos un cambio gramatical en cascada hacia el futuro).

En todos los casos, sin embargo, debemos tener cuidado de evitar la autocorrelación. Necesitamos seleccionar lenguas para los estudios tipológicos de manera que se evite el sobremuestreo de lenguas emparentadas y de lenguas en contacto areal.

# Universales Lingüísticos

# Qué son los Universales Lingüísticos?

- Los universales lingüísticos son patrones, principios o reglas que ocurren sistemáticamente en las lenguas naturales, pudiendo ser ciertos para todas ellas. Representan rasgos lingüísticos comunes compartidos entre las lenguas del mundo.
  - Revelan propiedades comunes del lenguaje humano, ofreciendo ideas sobre la facultad humana del lenguaje.
  - Pueden reflejar restricciones cognitivas, presiones funcionales (p. ej., eficiencia comunicativa) o propiedades inherentes de los sistemas lingüísticos (p. ej., la ley de Zipf).
  - El estudio de los universales es esencial para la teoría lingüística, ayudando a delinear los límites de la estructura o la gramática del lenguaje humano.

Podemos distinguir varios tipos de universales:

- **Universales Absolutos:**

- Estos patrones se aplican a toda lengua conocida y documentada sin excepción.
- Son relativamente pocos en número.
- *Ejemplos:*
  - "Todas las lenguas tienen sustantivos y verbos"
  - "Todas las lenguas tienen pronombres".
  - "Si una lengua es hablada (es decir, no una lengua de señas), tiene consonantes y vocales".

Podemos distinguir varios tipos de universales:

- **Universales Estadísticos (Tendencias):**

- Estas afirmaciones pueden no ser ciertas para todas las lenguas, pero ocurren con una frecuencia alta.
- Indican fuertes probabilidades o preferencias en las estructuras lingüísticas.
- *Ejemplos:*
  - "La gran mayoría de las lenguas tienen consonantes nasales" (p. ej., /m/, /n/).
  - "Con una frecuencia abrumadoramente mayor que la casualidad ("overwhelmingly greater-than-chance frequency"), las lenguas con un orden normal Sujeto-Objeto-Verbo (SOV) son posposicionales".



Podemos hacer más distinciones entre los universales absolutos y estadísticos:

- **Universales Implicacionales:**

- Se aplican a lenguas donde un rasgo particular (X) siempre o muy frecuentemente está acompañado por otro rasgo (Y).
- Toman la forma lógica: "Si una lengua tiene X, entonces también (o típicamente) tiene Y."
- Fueron investigados prominentemente por Joseph Greenberg.
- *Ejemplos:*
  - "Si una lengua tiene número gramatical trial (distinguiendo uno, dos, tres, muchos), también tiene número gramatical dual (distinguiendo uno, dos, muchos)".
  - Universal 27 de Greenberg: "Si una lengua tiene exclusivamente sufijos, tiene posposicionales. Si tiene exclusivamente prefijos, es tiene preposicionales".

Podemos hacer más distinciones entre los universales absolutos y estadísticos:

- **Universales No Implicacionales:**

- Simplemente afirman la existencia o no existencia de un rasgo particular, sin referencia a otros rasgos.
- Muchos universales absolutos son no implicacionales.
- *Ejemplo:* "Todas las lenguas tienen pronombres".

Los universales implicacionales pueden caracterizarse además por su direccionalidad:

- **Universales Bidireccionales:**

- En este tipo, dos rasgos se implican mutuamente, indicando un fuerte patrón de co-ocurrencia.
- La implicación funciona en ambas direcciones (Si X entonces Y, y Si Y entonces X).
- *Ejemplo:* "Las lenguas con posposiciones usualmente tienen un orden SOV (Sujeto-Objeto-Verbo), y asimismo, las lenguas SOV usualmente tienen posposiciones".
- Esto sugiere un vínculo sistémico entre el tipo de adposición y el orden básico verbo-objeto.

Los universales implicacionales pueden caracterizarse además por su direccionalidad:

- **Universales Unidireccionales:**

- Aquí, la implicación funciona en una sola dirección.
- Si X entonces Y, pero no necesariamente Si Y entonces X.
- *Ejemplo:* Las lenguas que colocan las cláusulas de relativo antes del sustantivo que modifican (orden RelN) usualmente tienen un orden SOV; por lo tanto, las cláusulas de relativo prenominales implican SOV.
- Sin embargo, el orden SOV no predice fuertemente las cláusulas de relativo prenominales, ya que las lenguas SOV muestran una ubicación variada de las cláusulas de relativo.

La existencia de diferentes tipos de universales subraya que la "universalidad" es un concepto matizado. Creo que podemos interpretar esto como que los universales tienen diferentes fuentes, que producen diferentes tipos de restricciones.

## Comparación Interlingüística y Espacio Tipológico

- Las lenguas exhiben una enorme diversidad superficial.
- **El Problema de la Comparabilidad (El "Problema de Manzanas y Naranjas"):**
  - Cómo pueden los lingüistas comparar significativamente estructuras gramaticales que se realizan de manera muy diferente entre las lenguas? Por ejemplo, lo que el inglés expresa con una preposición podría expresarse con un afijo de caso en latín, o un sustantivo relacional en kaqchikel (maya).
  - Comparar formas directamente suele ser bastante difícil debido a esta diversidad.

## La Solución de la Tipología: La Equivalencia Funcional como Punto de Partida:

- La comparación tipológica comienza identificando una **función** equivalente (significado, propósito comunicativo, papel semántico) entre las lenguas.
- Luego, examina las diversas **formas morfosintácticas** utilizadas para realizar esa función.
- Ejemplo: Para estudiar los causativos, uno identificaría las construcciones utilizadas para expresar "X causa que Y haga Z" en todas las lenguas, independientemente de si esto se hace mediante un afijo, un verbo auxiliar o una construcción de verbo serial.



- **Espacio Tipológico:** Se refiere al espacio conceptual multidimensional que abarca toda la variación estructural posible para un fenómeno o conjunto de fenómenos lingüísticos.
  - La tipología tiene como objetivo mapear este espacio, identificando:
    - Tipos atestiguados (aquellos encontrados en lenguas existentes).
    - Tipos no atestiguados pero lógicamente posibles (brechas sistemáticas).
    - Tipos comunes vs. raros.
  - La estructura de este espacio, incluyendo sus regiones "vacías", es altamente informativa sobre las restricciones en el lenguaje humano.
- Las **Jerarquías Implicacionales** mapean las dependencias dentro de este espacio. ej., si una lengua hace A, entonces hace B, y si una lengua hace B, hace C y E, etc.

# Tipología Sintáctica

# Introducción a la Tipología Sintáctica

- La tipología sintáctica estudia la variación interlingüística en la **estructura de la oración** y el **orden de palabras**.
  - **Orden Básico de Constituyentes:** El orden de Sujeto (S), Verbo (V), Objeto (O) y Objeto Indirecto (OI) en cláusulas declarativas neutras.
  - **Orden de Elementos dentro de Frases:**
    - Frases Nominales (p. ej., adjetivo-sustantivo, genitivo-sustantivo, numeral-sustantivo).
    - Frases Adposicionales (preposiciones vs. posposiciones).
  - **Relaciones Gramaticales:** Cómo se definen y marcan papeles como sujeto y objeto (vinculado al alineamiento morfosintáctico).
  - **Estrategias de Relativización:** Cómo se forman las cláusulas de relativo.
  - **Valencia y Voz:** Construcciones pasivas, causativas, applicativas y cómo se forman.
  - **Head-Marking vs. Dependent-Marking**
- El objetivo es identificar tipos comunes de orden de palabras, correlaciones entre diferentes parámetros de orden y restricciones universales en las estructuras sintácticas.

# Orden Básico de Constituyentes (Orden de Palabras)

- Se centra en el orden relativo de Sujeto (S), Verbo (V) y Objeto (O) en cláusulas principales transitivas y declarativas con argumentos de FN completos (p. ej., "El gato persiguió al ratón").
- Seis posibilidades lógicas: SOV, SVO, VSO, VOS, OVS, OSV.
- **Datos de WALS (Capítulo 81A):**
  - **Más Comunes:**
    - **SOV** (Sujeto-Objeto-Verbo): p. ej., japonés, turco, coreano, hindi, persa. (WALS: 564 lenguas)
    - **SVO** (Sujeto-Verbo-Objeto): p. ej., inglés, mandarín, español, suajili, ruso. (WALS: 488 lenguas)
  - **Menos Comunes:**
    - **VSO** (Verbo-Sujeto-Objeto): p. ej., galés, irlandés, tagalo, árabe clásico. (WALS: 95 lenguas)

# Orden Básico de Constituyentes (Orden de Palabras)

- Se centra en el orden relativo de Sujeto (S), Verbo (V) y Objeto (O) en cláusulas principales transitivas y declarativas con argumentos de FN completos (p. ej., "El gato persiguió al ratón").
- Seis posibilidades lógicas: SOV, SVO, VSO, VOS, OVS, OSV.
- **Datos de WALS (Capítulo 81A):**
  - **Raros:**
    - **VOS** (Verbo-Objeto-Sujeto): p. ej., malgache, fiyiano, tzotzil. (WALS: 25 lenguas)
    - **OVS** (Objeto-Verbo-Sujeto): p. ej., hixkaryana, apalaí. (WALS: 11 lenguas)
    - **OSV** (Objeto-Sujeto-Verbo): p. ej., nadëb, warao. (WALS: 4 lenguas)
  - Algunas lenguas no tienen orden dominante (**Orden de Palabras Flexible**), donde factores pragmáticos a menudo gobiernan la elección (p. ej., latín, ruso, nunggubuyu (Australia), muchas lenguas aborígenes australianas). (WALS: 189 lenguas "Sin un orden de palabras dominante").

Esta distribución resalta fuertes universales estadísticos (común S-inicial, O-final/medial) y rarezas (O-inicial muy raro).

# Correlaciones con el Orden Básico de Palabras (Universales de Greenberg)

El trabajo de Joseph Greenberg reveló fuertes correlaciones estadísticas entre el orden básico SOV/VO y otros órdenes de constituyentes:

- **Las lenguas OV (p. ej., SOV) tienden a mostrar:**
  - **Posposiciones** (Sustantivo Adposición): p. ej., japonés *Tōkyō ni* (Tokio a)
  - Orden **Genitivo-Sustantivo** (Poseedor Sustantivo): p. ej., turco *adamn evi* (del hombre casa)
  - Orden Adjetivo-Sustantivo (menos consistentemente)
  - El verbo auxiliar sigue al verbo principal
  - La cláusula de relativo precede al sustantivo (RelN)

# Correlaciones con el Orden Básico de Palabras (Universales de Greenberg)

## Las lenguas VO (p. ej., SVO, VSO) tienden a mostrar:

- **Preposiciones** (Adposición Sustantivo): p. ej., inglés *to Tokyo*
- Orden **Sustantivo-Genitivo** (Sustantivo Poseedor): p. ej., francés *la maison de l'homme* (la casa del hombre)
- Orden Sustantivo-Adjetivo (especialmente en VSO)
- El verbo auxiliar precede al verbo principal
- La cláusula de relativo sigue al sustantivo (NRel)

Estas correlaciones sugieren un impulso hacia la "armonía" o consistencia estructural en el orden núcleo-dependiente dentro de las lenguas.

- Las desviaciones (p. ej., el inglés es SVO pero tiene Adj-N) a menudo indican cambios históricos u otros factores complicadores.

# Correlaciones con el Orden Básico de Palabras (Universales de Greenberg)

Cómo podemos explorar la tipología de orden de palabras greenbergiana con nuestros corpus y Grew?



Las lenguas emplean varias estrategias para formar cláusulas de relativo (CR), que modifican a los sustantivos:

- **Orden de la Cláusula de Relativo y el sustantivo:**

- **sustantivo-Cláusula de Relativo (NRel):** La CR sigue al sustantivo núcleo.
  - Ejemplo: Español *el hombre [que vino ayer]*.
  - Este es el patrón dominante a nivel mundial.
- **Cláusula de Relativo-sustantivo (ReIN):** La CR precede al sustantivo núcleo.
  - Ejemplo: Japonés *[kinō kita] otoko no hito* ([ayer vino] hombre).
- Otros tipos: CR de núcleo interno (sustantivo núcleo dentro de la CR, p. ej., Cmique litom), CR correlativas (CR externa a la cláusula principal, vinculada anafóricamente, común en el sur de Asia), CR adjuntas, CR de doble núcleo.

## Estrategias para Vincular la FN Relativizada dentro de la CR (Keenan & Comrie):

- **Estrategia de Hueco (Estrategia Cero):** La posición de la FN relativizada dentro de la CR está vacía.
  - Ejemplo: Inglés *the book [I read \_\_\_\_]*. (el libro [que leí \_\_\_\_])
- **Estrategia de Pronombre Relativo:** Un pronombre especial (p. ej., *que, quien, el cual*) marca la posición relativizada o vincula la CR.
  - Ejemplo: Español *el hombre [quien se fue]*.
- **Estrategia de Retención de Pronombre (Pronombre "Resumtive"):** Un pronombre personal ordinario aparece en la posición relativizada dentro de la CR.

# Jerarquía de Accesibilidad de la Frase Nominal (Keenan & Comrie)

Keenan y Comrie (1977) propusieron una jerarquía universal que predice qué posiciones de FN son más fáciles de relativizar:

- **La Jerarquía:**

Sujeto > Objeto Directo > Objeto Indirecto > Oblicuo (Objeto Adposicional) > Genitivo (Poseedor) > Objeto de Comparativo

- **Implicación:** Si una lengua puede relativizar una posición en esta jerarquía usando una estrategia particular, también puede relativizar todas las posiciones más altas (a la izquierda) en la jerarquía usando esa misma estrategia (o una "más fácil").

# Jerarquía de Accesibilidad de la Frase Nominal (Keenan & Comrie)

## ● Ejemplos:

- El inglés puede relativizar todas las posiciones en la jerarquía.
- El malgache solo puede relativizar sujetos directamente.
- El euskera puede relativizar absolutivos, ergativos y objetos indirectos, pero no oblicuos o genitivos.
- Las lenguas pueden usar estrategias de "promoción" (p. ej., pasiva, antipasiva, applicativa) para hacer que las posiciones más bajas en la jerarquía sean accesibles a la relativización convirtiéndolas en sujetos u objetos directos.

# Jerarquía de Accesibilidad de la Frase Nominal (Keenan & Comrie)

Cómo podemos explorar la NPAH con nuestros corpus y Grew?

# Head-Marking vs. Dependent-Marking

Una distinción tipológica fundamental basada en dónde se marcan morfológicamente las relaciones gramaticales dentro de una frase: en el **núcleo** (head) o en el **dependiente** (dependent).

- **Lenguas de Head-Marking:** Los morfemas gramaticales que indican la relación aparecen en el núcleo de la frase.
  - *En la cláusula (Verbo como núcleo):* El verbo se marca para la concordancia con sus argumentos (dependientes).
    - Ejemplo (Tzutujil): *x-Ø-kee-k'aq* (COMP-3SG.ABS-3PL.ERG-tirar) 'ellos lo tiraron'. Marcadores de concordancia en el verbo.
  - *En la FN (sustantivo poseído como núcleo):* El sustantivo poseído se marca para la concordancia con el poseedor (dependiente).
    - Ejemplo (Tzutujil): *r-k'aajool n-b'esiino* (3ERG-hijo 1ERG-vecino) 'el hijo de mi vecino'. Prefijo en 'hijo' concuerda con 'vecino'.

# Head-Marking vs. Dependent-Marking

**Lenguas de Dependent-Marking:** Los morfemas gramaticales aparecen en los dependientes de la frase.

- *En la cláusula:* Los argumentos (dependientes) se marcan con caso para mostrar su relación con el verbo (núcleo).
  - Ejemplo (Checheno): *oox-a cha-ca txou-Ø qeelira* (1PL-ERG paja-INSTR techo-ABS cubrir.PST) 'Nosotros cubrimos el techo con paja'. Marcadores de caso en los argumentos.
- *En la FN:* El poseedor (dependiente) se marca con caso genitivo.
  - Ejemplo (Checheno): *loem-an k'orni* (león-GEN cachorro) 'el cachorro del león'. Caso genitivo en 'león'.
- **Otros Tipos:**
  - **Double-Marking:** Tanto el núcleo como los dependientes están marcados (p. ej., Mangarrayi, Jivaro).
  - **Zero-Marking (Sin Marcaje):** Ni el núcleo ni el dependiente están marcados abiertamente para la relación (p. ej., Pirahã).

# Head-Marking vs. Dependent-Marking

Cómo podemos explorar Head-Marking vs. Dependent-Marking con nuestros corpus y Grew? Probablemente necesitemos análisis morfológico...



## **Recopilación de Datos, Muestreo de Lenguas y Métodos Computacionales**

# Desafíos en la Recopilación de Datos y la Comparabilidad

Varios desafíos afectan la recopilación y el análisis de datos tipológicos:

- **Escasez de Datos y Peligro de Extinción de Lenguas:**

- Un gran número de las aproximadamente 7000 lenguas del mundo están poco documentadas, y muchas están en peligro de extinción, lo que limita los datos disponibles para un estudio tipológico exhaustivo.
- Esta escasez puede llevar a sesgos en las muestras y generalizaciones tipológicas si solo se consideran lenguas bien documentadas (a menudo lenguas grandes, coloniales o imperiales).

- **Categorías Tipológicas y Conceptos Comparativos:**

- Asegurar que categorías gramaticales como "sujeto", "adjetivo", "voz pasiva" o "tiempo" se definan de una manera que permita una comparación significativa entre lenguas estructuralmente diversas es un obstáculo importante, como ya hemos visto.

# Desafíos en la Recopilación de Datos y la Comparabilidad

Varios desafíos afectan la recopilación y el análisis de datos tipológicos:

- **Dependencia de Descripciones Existentes:**

- Los tipólogos dependen frecuentemente de gramáticas descriptivas escritas por otros lingüistas. La calidad, el detalle y el marco teórico de estas gramáticas pueden variar ampliamente, afectando potencialmente la precisión o la completitud de los datos tipológicos extraídos.

Tenga en cuenta, en general, que la fiabilidad de los hallazgos tipológicos depende en gran medida de la calidad, amplitud y comparabilidad de los datos lingüísticos primarios.

# Sesgo de Muestreo

El sesgo de muestreo puede distorsionar significativamente los hallazgos tipológicos si no se aborda adecuadamente:

- **Sesgo Genético:** La sobrerrepresentación de lenguas de unas pocas familias grandes (p. ej., indoeuropeas, níger-congo) puede llevar a que los rasgos comunes en esas familias se confundan con tendencias universales.
- **Sesgo Areal:** La sobrerrepresentación de lenguas de una región geográfica específica puede llevar a que los rasgos difundidos por contacto lingüístico (rasgos areales) se confundan con innovaciones independientes o tendencias universales.
- **Sesgo de Documentación:** Depender en gran medida de lenguas bien documentadas (a menudo lenguas nacionales importantes o aquellas con una extensa investigación lingüística) mientras se descuidan las lenguas poco documentadas o en peligro puede sesgar los resultados.

La práctica tipológica moderna enfatiza la creación de muestras grandes, genética y arealmente equilibradas para mitigar estos sesgos.

# Tipos de Muestreo de Lenguas

Dado que es imposible, y ni siquiera deseable, estudiar las 7000 lenguas, los tipólogos se basan en muestras.

- **Métodos de Muestreo (Rijkhoff & Bakker 1998):**

- **Muestreo Aleatorio:** Lenguas seleccionadas sin criterios específicos. Propenso a sesgos si el marco de muestreo está sesgado (p. ej., sobrerrepresentación de lenguas indoeuropeas) a menos que sea estratificado.
- **Muestreo Estratificado:** Crear una muestra que sea estadísticamente representativa del universo de muestreo controlando factores como la afiliación genética (familia lingüística) y el área geográfica. El objetivo es una muestra de lenguas en gran medida independientes, libre de sesgos genealógicos o areales.
- **Muestreo de Diversidad:** Maximizar la diversidad lingüística dentro de la muestra, incluyendo lenguas tipológicamente inusuales o "atípicas". Útil para probar la robustez de los universales propuestos y comprender toda la gama de variación, ya que "los tipos excepcionales ponen a prueba la regla".

## Bases de Datos Tipológicas

El auge de los métodos computacionales ha sido facilitado por el desarrollo de bases de datos tipológicas a gran escala.

- Típicamente bases de datos relacionales donde las lenguas están codificadas para varios rasgos tipológicos (p. ej., orden de palabras, marcaje de caso, tamaño del inventario de fonemas).
  - Permiten estudios estadísticos de grandes muestras sobre distribuciones y correlaciones de rasgos.
  - Facilitan la identificación de patrones areales y sesgos genealógicos.
  - Proporcionan datos para el modelado computacional de la estructura y evolución del lenguaje.

- **WALS (World Atlas of Language Structures):**
- <https://wals.info>
  - Una gran base de datos de propiedades estructurales (fonológicas, gramaticales, léxicas) (más de 2,662 lenguas)
  - Mapas que muestran la distribución geográfica de los valores de los rasgos, y capítulos de texto acompañantes
  - Muy bueno para patrones areales.



- **Glottolog:**

- <https://glottolog.org>

- catálogo de las lenguas, familias lingüísticas y dialectos del mundo, mantenido por el Instituto Max Planck.
- identificadores únicos (Glottocodes) para languoides, clasificaciones genealógicas (árbol de Glottolog) y extensas bibliografías de fuentes descriptivas.
- utilizado para el muestreo de lenguas, asegurando la diversidad genealógica.

- **PHOIBLE (Phonetics Information Base and Lexicon):**
- <https://phoible.org>
  - Un repositorio de datos de inventarios fonológicos interlingüísticos, con más de 3,020 inventarios de 2,186 lenguas distintas.
  - Datos extraídos de documentos fuente y compilados en un formato de búsqueda Unicode IPA.
  - Incluye datos de rasgos distintivos para los segmentos.

- **AUTOTYP:**

- <https://www.isle.uzh.ch/en/DLL/Databases-and-Methods/AUTOTYP.html>
  - Base de datos (1,125 lenguas / más de 2 millones de puntos de datos) centrada en dominios gramaticales específicos (p. ej., estructura de la FN, caso, concordancia).
  - "Autotipología (Bickel & Nichols 2002, Witzlack-Makarevich et al. 2021) significa que la mayoría de las variables no están predefinidas, sino que se desarrollan y refinan continuamente para mantener un nivel requerido de adecuación empírica."

Las herramientas computacionales facilitan el uso de estos grandes conjuntos de datos.

- El "package" de R '**lingtypology**' (George Moroz)
- <https://cran.r-project.org/web/packages/lingtypology/index.html>
- **Funcionalidad:**
  - Se conecta a la base de datos de Glottolog (Glottocodes, códigos ISO, afiliación, ubicación).
  - Crear mapas lingüísticos interactivos que muestran la distribución geográfica de los rasgos tipológicos.
  - Descargas de datos de otras bases de datos tipológicas importantes como WALS, AUTOTYP, PHOIBLE, etc.
  - Soporta Cross-Linguistic Linked Data (CLLD, <https://clld.org>):

- Cómo podemos explorar la tipología de orden de palabras greenbergiana con nuestros corpus, Grew y estas bases de datos?
- Cómo podemos explorar la NPAH con nuestros corpus, Grew y estas bases de datos?