



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE CÓMPUTO



Trabajo Terminal I

B - 042

Presenta:

- **Sánchez Mancilla Ivan Emiliano**
- **Juarez Solano Juan Martín**
- **Rojas Alarcón Sergio Ulises**

Directores de Trabajo Terminal:

- **M. en C. Elizabeth Moreno Galván**
- **M. en C. Jaime Arturo Lara Cázares**

Ciudad de México, 25 de Noviembre de 2024

Resumen

Abstract

Índice general

	Página
1. Introducción	6
1.1. Motivación	6
1.2. Problemática	6
1.3. Objetivos	7
1.3.1. Objetivo General	7
1.3.2. Objetivo Específicos	7
1.4. Alcance	8
1.4.1. Alcance Genral	8
1.4.2. Alcance Específico	8
1.4.3. Delimitación	8
1.5. Justificación	8
1.6. Propuesta de solución	9
1.7. Estado del arte	10
1.8. Metodología	11
2. Marco Teórico: conceptos teóricos	13
2.1. Comunicación	13
2.1.1. Tipos de Comunicación	13
2.1.2. Elementos de la comunicación	14
2.1.3. Barreras de la comunicación	15
2.2. Personas con discapacidad auditiva	16
2.2.1. Personas Oyentes	16
2.2.2. Personas con discapacidad auditiva (sordas)	16
2.2.3. Tipos de Discapacidad Auditiva	16
2.2.4. Brechas entre personas oyentes y personas con discapacidad auditiva	18
2.3. Lengua de Señas Mexicana	19
2.3.1. Definición de Lengua de Señas	19
2.3.2. Lengua de Señas Mexicana (LSM)	19
2.3.3. Dactilología	20
2.4. Inteligencia Artificial	20
2.4.1. Clasificación de la Inteligencia Artificial	21

2.5. Procesamiento de Lenguaje Natural (PLN)	23
2.5.1. Arquitectura de un sistema de PLN	23
2.5.2. Técnicas de PLN	24
2.5.3. Aplicaciones del PLN	29
2.6. MediaPipe	30
2.6.1. Herramientas de MediaPipe	30
2.6.2. MediaPipe Pose	31
2.7. Modelado de Animaciones 3D	34
2.7.1. Unity	34
2.8. Sistema Operativo Android	34
3. Desarrollo	36
3.1. Análisis y diseño de sistemas	36
3.1.1. Descripción del alcance	36
3.1.2. Entregables	36
3.1.3. Reglas de negocio	36
3.1.4. Requerimientos funcionales	37
3.1.5. Requerimientos no funcionales	37
3.1.6. Casos de uso	38
3.1.7. Diagrama de casos de uso	44
3.1.8. Mockups del sistema	44
3.1.9. Identificación de riesgos	44
3.1.10. Evaluación de riesgos	44
3.1.11. Mitigación de riesgos	44
3.2. Selección de los datos	44
3.2.1. Recolección de datos	44
3.3. Preparación de los datos	44
4. Conclusiones	45
5. Referencias	46
6. Trabajo relacionado	47
7. Desarrollo	48
8. Experimentos	49
9. Conclusiones y trabajo futuro	50

Capítulo 1

Introducción

1.1. Motivación

En México la comunicación es un derecho fundamental para todas las personas. Sin embargo, las personas de la comunidad sorda enfrentan barreras para poder comunicarse de manera efectiva. Si bien la Lengua de Señas Mexicana (LSM) es el principal medio de comunicación para las personas con discapacidad auditiva en México, en interacciones cotidianas o en situaciones de emergencia, las personas dependen de soluciones tecnológicas para poder comunicarse. Existen pocas herramientas digitales que permiten la traducción de español a LSM, pero en su mayoría no son precisas y muestran animaciones fragmentadas que dificultan la comprensión de los mensajes.

La motivación de este Trabajo Terminal es mejorar la accesibilidad de las personas sordas, a la par que se fomenta la inclusión social y tecnológica para esta comunidad, empleando técnicas de Inteligencia Artificial, Procesamiento de Lenguaje Natural y Modelado 3D para obtener una traducción más fluida y natural.

1.2. Problemática

La comunicación es una habilidad esencial que permite a los seres humanos poder interactuar con su entorno y compartir ideas a través de un sistema de símbolos [1]. No obstante, a menudo hay barreras en la comunicación que limitan el acceso a la información y dificultan la capacidad de las personas para expresarse y comunicarse [2].

A pesar de las barreras, las personas sordas han desarrollado su propio medio de comunicación que les permite superar las limitaciones fisiológicas y establecer una conexión efectiva con quienes comparten este lenguaje. En el caso de México, la Lengua de Señas Mexicana (LSM) es el principal medio de comunicación para las personas sordas en el país.

Se estima que en el territorio nacional hay un registro de 2.3 millones personas con esta discapacidad [3], lo que evidencia la magnitud de la población que depende de este lenguaje para interactuar en su entorno. Esta lengua posee su propia sintaxis, gramática, léxico y está compuesta por una combinación de señas, expresiones faciales y movimientos corporales que permiten transmitir ideas, mensajes, emociones y sentimientos [4].

1.3. Objetivos

1.3.1. Objetivo General

Desarrollar un prototipo de aplicación móvil en Android que, utilizando técnicas de Procesamiento de Lenguaje Natural y Modelado 3D, traduzca oraciones específicas empleadas en situaciones de emergencia, así como expresiones cotidianas como saludos y agradecimientos, del español a la Lengua de Señas Mexicana (LSM).

1.3.2. Objetivo Específicos

- Desarrollar el módulo de procesamiento de texto mediante técnicas de procesamiento de lenguaje natural, para interpretar oraciones específicas en español y transformarlas en sentencias manipulables para la traducción a la Lengua de Señas Mexicana (LSM).
- Construir un módulo de generación de animaciones 3D que represente visualmente las señas en Lengua de Señas Mexicana (LSM), a partir de las oraciones procesadas.
- Implementar el módulo de animación y transiciones fluidas utilizando Inteligencia Artificial, con el fin de optimizar significativamente la fluidez entre las señas, buscando una representación natural y continua del lenguaje de señas en los avatares.
- Crear una aplicación móvil en Android que integre los módulos de procesamiento de lenguaje natural, generación de avatares y transiciones fluidas entre los componentes del prototipo.
- Validar la funcionalidad y usabilidad de la aplicación móvil mediante pruebas con personas con discapacidad auditiva y personas oyentes, en escenarios simulados de emergencia, evaluando la precisión de la traducción y la fluidez de las animaciones.

1.4. Alcance

1.4.1. Alcance Genral

El TT consiste en el desarrollo de un prototipo de aplicación móvil que traduzca frases de español a LSM mediante animaciones 3D. La aplicación tendrá un conjunto predefinido de frases comunes para saludos y situaciones de emergencia, y en caso de no contar con una frase específica, se empleará la dactilología (representación de las letras de una palabra empleando las manos) para poder garantizar la comunicación.

1.4.2. Alcance Específico

Se hará uso de una interfaz de usuario simple e intuitiva, que sea de utilidad para la población objetivo. Mediante técnicas de Procesamiento Natural. Se pretende hacer un procesamiento del dataset que contiene las frases en español, para posteriormente realizar el modelado de las animaciones 3D empleando Mediapipe; dichas animaciones serán fluidas para mostrar una comunicación natural y fácil de entender.

1.4.3. Delimitación

Es importante aclarar que la traducción será únicamente de un canal de comunicación: de español a LSM. Lo anterior por la razón de que la traducción de LSM a español es más compleja por las diferencias estructurales y gramaticales entre ambos lenguajes.

1.5. Justificación

En México la comunicación inclusiva es un reto para las personas con discapacidad auditiva, debido a que la mayoría de la población no conoce la Lengua de Señas Mexicana (LSM), lo que condiciona a la comunidad sorda para poder expresar sus ideas y pensamientos, acceder a servicios esenciales o recibir apoyo en situaciones de emergencia. Pese a que ha habido avances tecnológicos a lo largo de los últimos años, las herramientas de traducción de español a LSM existentes presentan limitaciones, siendo las más importantes la falta de fluidez en las animaciones y la incompatibilidad entre versiones del sistema operativo de Android.

Este Trabajo Terminal busca desarrollar un prototipo de aplicación móvil que traduzca frases en español a LSM con animaciones en 3D fluidas, empleando técnicas de Procesamiento de Lenguaje Natural (PLN) y modelado 3D. Esto tiene el objetivo de contribuir a la accesibilidad, eliminar barreras de comunicación y fomentar la comunicación inclusiva entre personas oyentes y personas con discapacidad auditiva.

Se espera que este desarrollo sirva de referencia para sentar las bases de futuras investigaciones relacionadas a la traducción e interpretación de lenguas de señas empleando técnicas de Inteligencia Artificial (IA).

1.6. Propuesta de solución

Este proyecto consiste en desarrollar un prototipo que aborde estos problemas, centrándose en la fluidez y precisión de las animaciones de LSM, mediante Inteligencia Artificial, y apoye la interacción entre personas oyentes y personas con discapacidad auditiva mediante Lengua de Señas Mexicanas (LSM), empleando técnicas de procesamiento de lenguaje natural y modelado 3D. El traductor estará enfocado en un conjunto limitado de frases predefinidas, como expresiones de emergencia, saludos y agradecimientos. En caso de no encontrar la frase deseada, el sistema procederá a deletrear la palabra o frase mediante el alfabeto de la Lengua de Señas Mexicanas (LSM), asegurando siempre una respuesta en la comunicación.

El prototipo estará diseñado para la última versión de Android, debido a que este es el sistema operativo más utilizado en México [5], por lo que la solución propuesta tendrá un mayor alcance y beneficiará a una mayor cantidad de usuarios. Según diversas estadísticas [5], la penetración de dispositivos Android supera con creces a la de iOS.

Para el desarrollo del prototipo se utilizará Media Pipe [6], una biblioteca eficiente para procesamiento de gestos y movimientos corporales, ideal para capturar de manera precisa los gestos de la Lengua de Señas Mexicana. Esto permitirá obtener datos precisos de las señas, garantizando una traducción más confiable y fluida, mientras se mantiene un rendimiento óptimo incluso en dispositivos de gama media.

Además, el uso de aplicaciones de Procesamiento de Lenguaje Natural (PLN) es fundamental para el procesamiento del español a LSM, ya que permitirá que el sistema no solo reconozca palabras aisladas, sino también frases y oraciones completas, adaptando la traducción al contexto y mejorando la comunicación en situaciones más complejas.

Por otro lado, se busca crear representaciones visuales en 3D para poder facilitar la visualización de señas. Las personas con discapacidad auditiva podrán visualizar una animación asociada a una de las frases que están incluidas dentro del prototipo.

Productos Esperados

- Dataset normalizado de LSM.
- Set de animaciones con avatares 3D fluidos e interactivos.
- Aplicación móvil en Android.
- Documentación del sistema.

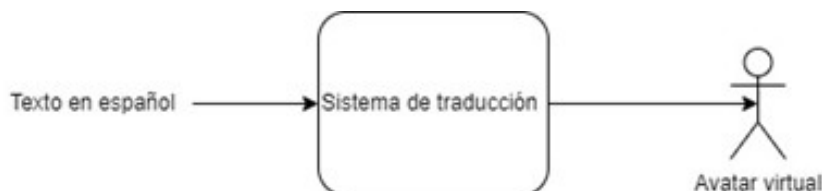


Figura 1.1: Diagrama del funcionamiento de la App, elaboración propia.

1.7. Estado del arte

Se han desarrollado diversas investigaciones y desarrollos de sistemas de software que abordan la problemática de la comunicación entre personas oyentes y no oyentes. A continuación se describen los trabajos relacionados:

- Translation of Spanish Text to Mexican Sign Language Glossed Text Using Rules and Deep Learning. En este trabajo se presenta una arquitectura para traducir de español a Lengua de Señas Mexicana (LSM), cuyos resultados fueron evaluados con las métricas BLEU y WER. En este artículo se encontró que la traducción con técnicas tradicionales tiene un mejor desempeño que el aprendizaje profundo [7].
- Resource Creation for Automatic Translation System from Texts in Spanish into Mexican Sign Language. Este artículo presenta la creación de recursos lingüísticos para la traducción automática del español escrito a LSM, así como el sistema que la implementa. En colaboración con la Casa de la Cultura de los Sordos (CDMX), se tradujeron 150 oraciones pertenecientes a 13 estructuras gramaticales reconocidas por el sistema, identificando 100 signos que pueden representar una oración completa. Se observó que, ante la falta de un signo específico, las personas sordas tienden a deletrear la palabra. El sistema traduce de forma literal al buscar y reproducir palabras en su base de datos, además de contar con signos que representan expresiones completas [8].
- Aplicación “Voz y Señas”. Esta aplicación, desarrollada por el Instituto de Pedagogía en conjunto con TecnoPrótesis y Bienestar Incluyente A.C., permite traducir la LSM por medio del habla o escribiendo texto. Sus objetivos son favorecer la comunicación entre una persona sorda y una persona oyente,

y ser una herramienta auxiliar en los procesos de alfabetización, redacción de textos y comprensión lectora [9].

- Hetah. Servicio en línea que funciona como traductor de Lengua de Señas Colombiana (LSC), que permite la comunicación entre personas oyentes y no oyentes mediante un Avatar 3D. Se debe ingresar la frase que se desea traducir y posteriormente, el avatar realizará la traducción mediante gestos [10].
- SignAloud. Sistema de Inteligencia Artificial que emplea guantes capaces de reconocer gestos de las manos correspondientes a palabras y frases en Lenguaje de Señas Americana. Dichos guantes contienen sensores que registran la posición y el movimiento de las manos, para generar datos que son enviados de forma inalámbrica a una computadora central. La computadora analiza los datos de los gestos empleando redes neuronales, y si los datos coinciden con un gesto, la palabra o frase asociada se pronuncia en un altavoz [11].
- Sistema traductor de la Lengua de Señas Mexicana mediante dactilología y de español a español signado. Sistema de Inteligencia Artificial capaz de ayudar a entablar un diálogo entre una persona sorda y otra oyente, mediante la traducción de las señas de dactilología a texto plano en español y de forma análoga, la traducción del texto plano a español a español signado [12].

1.8. Metodología

El proyecto se desarrollará bajo la metodología Scrum, un marco de trabajo ágil que organiza la colaboración del equipo a través de roles, artefactos y reglas que garantizan su correcta implementación [13][14]. Uno de sus principios clave es la configuración de equipos autogestionados y multifuncionales, lo que permite tomar decisiones autónomas sin depender de directrices externas [14].

La estructura de Scrum se basa en ciclos iterativos llamados Sprints, en los cuales se genera un incremento funcional del producto. Cada Sprint se gestiona como un proyecto independiente con objetivos específicos [14]. Además, estos ciclos incluyen cinco elementos clave: reunión de planeación, Daily Scrum, trabajo de desarrollo, revisión del Sprint y retrospectiva del Sprint, asegurando así una mejora continua en cada iteración. Para este desarrollo, los Sprints tendrán una duración de entre dos y cuatro semanas, lo que facilitará una transición progresiva hacia esta metodología y garantizará un avance constante.

Otro punto a destacar es que Scrum define tres roles esenciales. En primer lugar, el Scrum Master guía la implementación de la metodología y facilita la resolución de impedimentos sin gestionar directamente el desarrollo [15]. En segundo lugar, el Product Owner representa a los interesados y gestiona el Product

Backlog, priorizando las funcionalidades para maximizar el valor del producto [15]. Por último, el equipo de desarrollo transforma estos requerimientos en incrementos funcionales, operando sin jerarquías y con un tamaño ideal de tres a nueve integrantes [15].

En este proyecto, el equipo asumirá exclusivamente el rol de desarrolladores, encargándose de transformar los requerimientos en incrementos funcionales del producto. La estructura del equipo se distribuirá en tres áreas especializadas. El desarrollador de animaciones 3D y MediaPipe diseñará avatares, implementará gestos y optimizará las transiciones. El desarrollador de Android estructurará la aplicación, desarrollará interfaces y conectará los módulos. Por su parte, el desarrollador de PLN implementará el procesamiento de lenguaje natural, adaptará las frases al contexto de LSM y optimizará la comunicación con los módulos de animación.

Esta distribución permitirá aplicar Scrum de manera eficiente, asegurando un desarrollo iterativo y coordinado, en el que cada integrante contribuirá activamente al avance del proyecto mediante la integración de sus respectivas áreas.



Figura 1.2: Metodología Scrum, obtenido de [1].

Capítulo 2

Marco Teórico: conceptos teóricos

2.1. Comunicación

La comunicación se trata de un proceso dinámico, en el que necesariamente participan una fuente o emisor que envía un mensaje a través de un canal o medio a un potencial receptor que, a su vez, puede convertirse también en emisor [16]. Una comunicación efectiva se logra cuando se transmite el mensaje de una forma clara y efectiva para el receptor sin generar dudas ni confusiones [17].

Comunicar es el acto que permite establecer relaciones efectivas, compartir experiencias, experimentar emociones y sentimientos, así como hacer que los demás lo experimenten [18].

2.1.1. Tipos de Comunicación

Uno de los tipos de comunicación está basado en si se usan palabras o no, es decir, comunicación verbal o no verbal [20]:

- Comunicación verbal: se emplean palabras, y se lleva a cabo a través del habla o de manera escrita.
- Comunicación no verbal: se emplea el lenguaje corporal, gestos, signos no lingüísticos y sonidos que no forman palabras.

Otro de los tipos de comunicación son la formal y la informal, las cuales se describen a continuación:

- Formal: se utiliza un lenguaje especializado y estandarizado, sin errores ni coloquialismos, además de que se toman en cuenta las jerarquías sociales. Se basan en hechos y datos concretos.
- Informal: no se emplea lenguaje estandarizado, no se siguen protocolos jerárquicos y se emplean coloquialismos.

Un tercer tipo de clasificación es basada al tipo de acto comunicativo, en el que se enlistan los siguientes elementos [20]:

- Comunicación intrapersonal: conversaciones que un ser humano entabla consigo mismo.
- Comunicación interpersonal: intercambio de ideas y pensamientos entre dos personas, la cuál debe ser directa e interactiva.
- Comunicación grupal: intercambio de ideas y pensamientos entre un grupo de personas, las cuales se comunican con un propósito.
- Comunicación masiva: comunicación dirigida a un gran número de personas, mediante un medio masivo de comunicación.

2.1.2. Elementos de la comunicación

En el proceso de comunicación hay una serie de elementos que hacen posible la transmisión de un mensaje. A continuación, se enlistan cada uno de ellos:

- Emisor. Dentro del proceso comunicativo, es el individuo que inicia el intercambio de información al transmitir el mensaje [18]. Dicho mensaje debe ser codificado en un sistema de símbolos que deberá ser entendible para el receptor.
- Receptor. Individuo que recibe el mensaje enviado, el cual es interpretado con base en las experiencias, opiniones, contexto y situación del receptor [16]. El receptor también puede ser el emisor.
- Código. Sistema de signos que es empleado tanto por el emisor como por el receptor para llevar a cabo el proceso de comunicación. Ese sistema debe ser conocido por ambos para facilitar la codificación y decodificación [20].
- Mensaje. Es la información que el emisor transmite al receptor por medio del código [19].
- Canal. El canal es el medio en el que los mensajes del emisor se transmiten hacia el receptor [16].
- Contexto. Se refiere a la situación en la que se lleva a cabo el proceso de comunicación, y la cual tiene influencia directa en el entendimiento e interpretación del mensaje [19].
- Ruido o interferencia. En el proceso de comunicación puede haber factores externos que dificultan o impiden el entendimiento de los mensajes [20].

Como se mencionó anteriormente, la comunicación es un proceso indispensable para la interacción humana ya que por medio de ella las personas pueden intercambiar ideas, pensamientos y emociones. No obstante, como se menciona en el concepto de ruido, en ocasiones hay elementos que impiden que la comunicación se lleve a cabo, como lo pueden ser las barreras de la comunicación.

2.1.3. Barreras de la comunicación

Las barreras de la comunicación son elementos que limitan o dificultan que las personas puedan comunicarse, a la par que se dificulta su proceso de comunicación [2]. Son todas las perturbaciones que sufre un mensaje, en cualquiera de los elementos que forman parte del proceso de comunicación.

Los principales tipos de barreras son:

1. **Barreras físicas.** Son interferencias causadas por elementos del entorno o en el medio donde se lleva a cabo la comunicación [21].
2. **Barreras psicológicas.** Son aquellas que surgen por emociones, prejuicios o estados mentales que afectan la interpretación del mensaje [21].
3. **Barreras semánticas.** Surgen cuando hay confusión en el significado de las palabras, debido a una interpretación incorrecta del lenguaje. Generalmente ocurren cuando se habla en un idioma que el emisor o el receptor no entienden, o se emplean conceptos técnicos desconocidos [21].
4. **Barreras administrativas.** Generalmente se presentan en entornos laborales, son causadas por falta de planeación, malentendidos, falta de claridad en los procesos de comunicación y distorsiones semánticas [21].
5. **Barreras culturales.** Este tipo de barreras se presentan cuando hay diferencias en costumbres, valores, normas o expresiones entre culturas, que imposibilitan la comunicación [21].
6. **Barreras interpersonales.** Hace referencia a las barreras en las que hay suposiciones incorrectas y diferentes percepciones [21].
7. **Barreras tecnológicas.** Fallas y limitaciones que se presentan en medios tecnológicos empleados para la comunicación [21].
8. **Barreras fisiológicas.** Impedimentos físicos o biológicos causados por deficiencias en los sentidos, enfermedades o condiciones médicas que afectan cualquiera de los sentidos de manera parcial o total, afectando la transmisión de información [21]. Por ejemplo, voz débil, pronunciación defectuosa, sordera, problemas del habla, problemas visuales, etc.

Para efectos de este Trabajo Terminal se analizarán las barreras fisiológicas, concretamente las que son causadas por problemas de sordera. En el siguiente apartado se describen los términos correctos para referirse a las personas con capacidad de escucha y a las personas con discapacidad auditiva.

2.2. Personas con discapacidad auditiva

2.2.1. Personas Oyentes

Un oyente es una persona con la capacidad de escuchar sonidos que le permiten interpretar mensajes. El término procede del verbo oír, que refiere a la capacidad que posee un individuo para poder percibir sonidos [22].

2.2.2. Personas con discapacidad auditiva (sordas)

La discapacidad auditiva se define como la pérdida de la función del sistema auditivo, teniendo como consecuencia una discapacidad para poder oír, lo que dificulta el acceso al lenguaje oral [s1].

De acuerdo con la Federación Mundial de Sordos, existen aproximadamente 70 millones de personas sordas en todo el mundo, las cuales emplean más de 300 diferentes lenguas de señas [s2]. Las lenguas de señas varían entre países, presentando cambios principalmente en la estructura gramatical, sintaxis, vocabulario, signos, alfabeto y expresiones corporales [22].

Por otro lado, de acuerdo con la Secretaría de Salud [3], en México hay aproximadamente 2.3 millones de personas con discapacidad auditiva, de las cuáles más del 50 % son mayores de 60 años, 34 % tienen entre 30 y 59 años, y el 2 % son niñas y niños.

Las principales causas de problemas de audición son antecedentes familiares de sordera heredados, edad avanzada, enfermedades infecciosas, exposición continua a sonidos intensos, entre otras [3].

Las personas sordas enfrentan consecuencias en ámbitos académicos, laborales, sociales y emocionales, debido a que las situaciones de aislamiento, deficiencia en la comunicación y dificultades del día a día repercuten negativamente para integrarse en grupos y para socializar [s21].

2.2.3. Tipos de Discapacidad Auditiva

La discapacidad auditiva se clasifica en tres tipos según distintos criterios: según la parte del oído afectada, según el grado de pérdida auditiva y según el

momento en que se adquiere [s22]:

Según la parte del oído afectada

- **Hipoacusia conductiva.** Es producida por un impedimento en el trayecto de las ondas sonoras del oído externo y medio al oído interno, causado por tumores, perforación del tímpano, traumatismos o malfunciones del oído.
- **Hipoacusia neurosensorial.** Se produce cuando el nervio auditivo o las células ciliadas son dañadas, ya sea por herencia, anormalidades al momento del nacimiento, exposición a ruidos fuertes, traumatismos, entre otras causas.
- **Hipoacusia mixta.** Combinación de hipoacusia conductiva e hipoacusia neurosensorial, causadas por anormalidades al nacer, infecciones, tumores y lesiones en la cabeza.

Según el grado de pérdida

Una audición normal está en el rango de los 0-20 dB. A partir de ese rango, se establece la siguiente clasificación de acuerdo con los dB que se hayan perdido:

- **Leve:** 20-40 dB.
- **Moderada:** 40-70 dB.
- **Severa:** 70-90 dB.
- **Profunda:** más de 90 dB.

Según el momento de la adquisición

En esta clasificación, la discapacidad auditiva puede ser:

- **Hereditaria.** La discapacidad está contenida en algunos de los genes de uno o ambos progenitores.
- **Adquirida.** La discapacidad puede ser prenatal (antes del nacimiento) o postnatal (después del nacimiento), y en este último caso se deben tomar en cuenta otros criterios:
 - **Prelocutiva:** antes del desarrollo del lenguaje.
 - **Postlocutiva:** después del desarrollo del lenguaje.

2.2.4. Brechas entre personas oyentes y personas con discapacidad auditiva

En el plano sociocultural el lenguaje es esencial en las formas de comunicación en una comunidad, pero cuando no todos los individuos pueden responder a esa lógica comunicativa se crean brechas en los discursos que giran en torno a las formas de relacionarse con los demás, puesto que aquellos que tienen códigos y configuraciones diferentes pasan a estar en un plano de invisibilidad [s24].

La comunidad sorda, a pesar de ser un grupo portador de un lenguaje cultural particular, debe responder a la lengua “natural” de las personas oyentes, y de no poder hacerlo ocasiona que sean excluidos en diferentes escenarios de la vida cotidiana. Esta comunidad ha sido estereotipada como personas incapaces o con limitaciones para insertarse en la sociedad, por lo que si no pueden entrar en la “lógica natural” para comunicarse con las personas, se ven forzados a interactuar solamente con las personas que comparten su misma condición [s24].

A lo largo de los últimos años, se han realizado múltiples esfuerzos a nivel gubernamental y se han puesto en marcha discursos que giran alrededor del reconocimiento e inclusión de todas las personas por igual, para garantizar una mayor participación de las personas con discapacidad auditiva en escenarios sociales [s24].

No obstante, lo expresado en la legalidad dista mucho de las realidades particulares de las personas sordas en el marco sociocultural. La comunidad sorda ha sido reconocida como minoría lingüística, y por sus mismas condiciones, ha sido ubicada socialmente en el plano de la exclusión y la invisibilidad [s24].

La presencia de barreras de comunicación generan aislamiento e impiden el desarrollo de una existencia satisfactoria, lo que puede generar graves problemas psicológicos como la depresión, ansiedad, insomnio, estrés, ideas paranoides y sensibilidad interpersonal [s1].

Además, la comunidad sorda presenta dificultad para acceder a la información proveniente de la televisión, radio, llamadas telefónicas, megafonías en estaciones de metro y salidas de aeropuertos, etc., debido a que esta es principalmente transmitida hacia la población oyente.

A pesar de que las personas sordas presentan muchas dificultades en su vida diaria, hoy en día disponen de numerosas herramientas de apoyo para impulsar su inclusión en entornos sociales y favorecer su crecimiento personal, como lo son las prótesis auditivas, señales acústicas y su propia Lengua de Señas. En este Trabajo Terminal, únicamente se centrará el estudio en las Lenguas de Señas, concretamente en la Lengua de Señas Mexicana (LSM).

2.3. Lengua de Señas Mexicana

2.3.1. Definición de Lengua de Señas

El lenguaje de señas es definido como la lengua natural de expresión y configuración gesto-espacial y percepción visual gracias a la cual los sordos pueden comunicarse con su entorno social. Está basado en movimientos y expresiones a través de manos, ojos, rostro, boca y cuerpo [s25].

En el mundo existen cerca de 300 lenguajes de señas distintos, siendo así que cada país posee su propia lengua de señas. Por ejemplo, la Lengua de Señas Mexicana (LSM) no es igual a la Lengua de Señas Española (LSE), que a pesar de estar articulados en el mismo idioma (español), no comparten muchos signos en común, debido a que ambas lenguas presentan señas que pueden ser regionalismos de cada país [s25].

Por su parte, la Lengua de Señas Mexicana (LSM) cuenta con su propio vocabulario y gramática, y es considerado como un lenguaje por cuenta propia, debido a que es completamente capaz de expresar una amplia gama de pensamientos y emociones como cualquier otra lengua [s25].

2.3.2. Lengua de Señas Mexicana (LSM)

La Ley General para la Inclusión de las Personas con Discapacidad [s26] define a la Lengua de Señas Mexicana (LSM), en el Artículo 2, como la lengua de una comunidad de sordos que consiste en una serie de signos gestuales articulados con las manos y acompañados de expresiones faciales, mirada intencional y movimiento corporal, dotados de función lingüística, forma parte del patrimonio lingüístico de dicha comunidad y es tan rica y compleja en gramática y vocabulario como cualquier lengua oral [s26]

Por su parte, el Artículo 20 de dicha ley establece que los medios de comunicación implementarán el uso de tecnología, más concretamente, de intérpretes de LSM que permitan a la comunidad de sordos las facilidades de comunicación y el acceso al contenido de su programación [s26]

En México hay entre 87,000 y 100,000 personas hablantes de LSM que la dominan y la emplean como vía de comunicación, siendo incluso una población mucho

más grande que algunas comunidades hablantes de lenguas indígenas del país [s27].

2.3.3. Dactilología

La dactilología es un sistema de comunicación que transmite información mediante el deletreo manual, y en ocasiones es usado en conjunto con la lengua de signos. Se emplea la mano de diferente manera para pronunciar cada una de las letras [s22].

Otra definición de la dactilología es que es la representación manual de cada una de las letras que componen el alfabeto, para poder transmitir a las personas sordas cualquier palabra que se desee comunicar. Todas las lenguas de señas poseen mecanismos internos que les permiten generar mensajes [s23].

Para comunicarse por medio de dactilología se emplea la mano dominante a la altura de la barbilla, en conjunto con la articulación oral, siendo necesario que la cara y la boca sean visibles [s23]. Principalmente se usa para sustantivos, nombres propios, direcciones y palabras para los cuales no existe un signo creado.

Si bien la discapacidad auditiva representa una barrera de la comunicación, las personas sordas en los últimos años han buscado superar esa barrera con ayuda de dispositivos tecnológicos que puedan fungir como intérpretes. El desarrollo de la Inteligencia Artificial (IA), más concretamente las técnicas de Procesamiento de Lenguaje Natural (PLN) y modelado de animaciones 3D, han ayudado a crear nuevos sistemas que faciliten la interacción entre personas oyentes y personas de la comunidad sorda, derribando las barreras de la comunicación.

2.4. Inteligencia Artificial

La Inteligencia Artificial (IA) es la capacidad que poseen las máquinas para usar algoritmos y aprender de los datos para tomar decisiones tal como lo haría un ser humano [s3]. A diferencia del ser humano, la IA no necesita descansar y es capaz de analizar grandes cantidades de información, reduciendo el margen de error.

La IA se basa en el uso de algoritmos y tecnologías de aprendizaje automático para dar a las máquinas la capacidad de aplicar ciertas habilidades cognitivas y realizar tareas por sí mismas de manera autónoma o semiautónoma. A medida que la IA mejora, muchos procesos son más eficientes y algunas tareas que parecían complicadas se realizan con mayor rapidez y precisión [s31].

2.4.1. Clasificación de la Inteligencia Artificial

La IA puede ser clasificada de varias maneras, ya sea a partir de su grado de capacidad cognitiva o a partir de su grado de autonomía [s31].

Clasificación a partir de su grado de capacidad cognitiva:

- Inteligencia Artificial débil o limitada. Está diseñada para realizar tareas específicas de manera eficiente, pero no tiene la capacidad de razonar de manera generalizada ni de aprender de nuevas situaciones [s31].
- Inteligencia Artificial general o fuerte (AGI). Este tipo de IA tiene la capacidad de realizar varias tareas cognitivas como el razonamiento, el aprendizaje y la resolución de problemas. A diferencia de la IA débil, la IA fuerte es capaz de adaptarse a nuevas situaciones y entornos [s31].
- Super Inteligencia Artificial (ASI). Tiene la capacidad de realizar prácticamente cualquier tarea compleja que requiere Inteligencia Humana, ya que es muy poderosa, y puede superar a los seres humanos en términos de capacidad cognitiva y de aprendizaje [s31].

Clasificación de acuerdo con su grado de autonomía:

- Inteligencia Artificial Reactiva. Este tipo de IA realiza tareas específicas de manera autónoma, pero no tiene la capacidad de recordar eventos pasados ni de anticipar situaciones futuras. Es útil en situaciones en las que se requieren respuestas rápidas y precisas a situaciones específicas [s31].
- Inteligencia Artificial Deliberativa. Tiene la capacidad de planificar y tomar decisiones basándose en información del entorno y en objetivos predeterminados. Es decir, puede analizar situaciones y elegir opciones que le permitan cumplir con objetivos, o adaptarse a entornos empleando información del pasado y del futuro [s31].
- Inteligencia Artificial Cognitiva. Se caracteriza por su capacidad de imitar las funciones cognitivas humanas como lo son el razonamiento, la percepción y el aprendizaje, y tienen la capacidad de adaptarse a nuevas situaciones y entornos [s31].
- Inteligencia Artificial Autónoma. Es capaz de interactuar de manera autónoma con su entorno, tomar decisiones y aprender de nuevas situaciones, y

cambiar sus objetivos y estrategias en función de las estrategias sin la necesidad de la intervención humana [s31].

La Inteligencia Artificial emplea diferentes técnicas, las cuales se enlistan a continuación [s4]:

- **Búsqueda de soluciones.** Tiene por objetivo encontrar mecanismos de deducción y búsqueda de soluciones para la resolución de problemas cuando no se cuenta con un método directo [s4].
- **Representación del conocimiento.** Elaboración de métodos y técnicas eficientes que sean capaces de organizar conocimientos en un sistema, para posteriormente ser usados en la búsqueda de soluciones para diferentes problemáticas [s4].
- **Reconocimiento de patrones.** Son técnicas de clasificación para identificar subgrupos midiendo el parecido o similitud entre formas, con el objetivo de obtener conclusiones [s4].
- **Robótica.** Esta técnica tiene por objetivo la construcción de robots inteligentes capaces de funcionar con autonomía, que sean capaces de realizar procesos mecánicos y manuales con el fin de obtener mayor productividad, suplir mano de obra y proporcionar flexibilidad en procesos industriales [s4].
- **Redes Neuronales.** Son sistemas compuestos por estructuras de red, con un gran número de conexiones entre diferentes capas de procesadores, que a su vez tienen asignadas diferentes funciones. Las redes neuronales efectúan una labor de aprendizaje por la reproducción de las salidas de un conjunto de entrenamiento [s4].
- **Algoritmos genéticos.** Son los tipos de algoritmos que tratan de emular el proceso de selección natural a un problema dado, en el que se aplican operadores genéticos para evaluar cada una de las soluciones propuestas. Se emplean procedimientos de búsqueda y optimización para mejorar las soluciones existentes y generar nuevas [s4].
- **Sistemas expertos.** Sistemas que almacenan conocimientos de expertos sobre un área o campo especializado, para obtener una solución mediante una

deducción lógica. Se incluyen juicios, experiencias, opiniones, reglas de evaluación para proveer asesoría sobre diversas tareas [s4].

- **Procesamiento de Lenguaje Natural (PLN).** El Procesamiento de Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial (IA) centrada en el diseño de métodos y algoritmos que toman como entrada o producen como salida datos en la forma del lenguaje humano, ya sea en forma de texto, audio o animación [s41].

En este trabajo nos centraremos en la técnica de Procesamiento de Lenguaje Natural (PLN). En el siguiente apartado se profundizará más en el concepto, características y usos del PLN.

2.5. Procesamiento de Lenguaje Natural (PLN)

El Procesamiento de Lenguaje Natural (PLN, o NLP por sus siglas en inglés) es el campo de estudio que busca entender cómo funciona el lenguaje, su construcción, la generación de nuevo lenguaje, así como todas las tareas que tienen relación con el tratamiento del lenguaje como lo es la generación de texto, traductores, generadores de resúmenes, chatbots, entre otros [s44].

El PLN emplea el lenguaje natural para establecer comunicación entre un ser humano y una computadora. Esta última deberá entender las oraciones que le sean proporcionadas mediante modelos que le ayuden a entender los mecanismos humanos relacionados con el lenguaje [s43].

2.5.1. Arquitectura de un sistema de PLN

La arquitectura de un sistema de PLN está dividida en niveles:

- a. **Nivel Fonológico.** Explica cómo es que las palabras se relacionan con los sonidos que representan.
- b. **Nivel Morfológico.** Indica cómo es que las palabras se construyen a partir de unidades de significado más pequeñas llamadas morfemas.
- c. **Nivel Sintáctico.** Se analiza cómo es que las palabras se unen para formar oraciones, entendiendo la función estructural que cada palabra posee.
- d. **Nivel Semántico.** Se refiere al significado de las palabras, y cómo los mismos se unen para darle sentido a una oración, considerando también el contexto de la oración.
- e. **Nivel Pragmático.** Trata de cómo las oraciones son empleadas en diferentes situaciones y cómo es que el uso afecta el significado de las mismas.



Figura 2.1: Arquitectura de un Sistema de Procesamiento de Lenguaje Natural, obtenido de [s43].

Los pasos que sigue la arquitectura del sistema de PLN es la siguiente [s43]:

1. El usuario le expresa a la computadora lo que desea hacer.
2. La computadora analiza las oraciones que el usuario le proporciona, en el sentido morfológico y sintáctico. Es decir, se verifican los componentes léxicos definidos y se verifica si se cumple un orden gramatical entre los elementos identificados.
3. Posteriormente, se realiza un análisis sintáctico de las oraciones, para saber cuál es el significado de cada oración.
4. Después de realizar el paso anterior, se lleva a cabo un análisis pragmático de todas las oraciones juntas. Al final de este paso, la computadora obtiene la expresión final.
5. Una vez obtenida la expresión final, la misma es ejecutada para obtener un resultado que será proporcionado al usuario.

2.5.2. Técnicas de PLN

El PLN se apoya de un conjunto de técnicas mediante las cuales se extrae información determinada de un texto. A continuación, se describen algunas de las técnicas más comunes utilizadas:

1. Detección de oraciones. La detección de oraciones es una de las técnicas básicas correspondientes al nivel de procesamiento sintáctico, que funciona recortando una secuencia de caracteres entre dos signos de puntuación; el signo debe estar acompañado por un espacio en blanco y se excluye el caso de la primer frase, y en posibles ocasiones la última frase.

La detección de oraciones puede presentar algunas dificultades a la hora de procesar títulos, abreviaturas, o algunos elementos que no siguen algún patrón de texto plano. En el caso de las abreviaturas, se utilizan palabras cargadas en el modelo, empleando un modelo por idioma, ya que el mismo posee los símbolos o abreviaturas necesarias para detectar las sentencias [s46].

Pierre Vinken, 61 años, se unirá a la junta como director no ejecutivo 29 noviembre. El Sr. Vinken es presidente de Elsevier NV, el grupo editorial holandés.	<pre><Párrafo inicio="0" fin="156"> <Oración inicio="0" fin="83" /> <Oración inicio="84" fin="156" /></pre>
---	---

Figura 2.2: Ejemplo de la delimitación de oraciones dentro de un párrafo, obtenido de [s46].

El ejemplo anterior muestra que el modelo en español determina que “Sr.” es una abreviatura de la palabra “Señor”, y por consiguiente ignora el signo de puntuación como final de la oración.

2. Segmentación por palabras. Después de que se identificaron cada una de las oraciones que componen el texto, se procede a la segmentación por palabras, más conocida como analizador léxico o “Tokenizer”. Esta técnica pertenece al nivel léxico y consiste en la identificación de tokens, que son unidades lingüísticas como palabras, puntuación, números, caracteres alfanuméricos, etc. Para identificar tokens en idiomas modernos, se delimitan espacios en blanco con límites de palabra, entre comillas, paréntesis y puntuación. El trato con las abreviaciones es similar a la detección de oraciones, ya que no existen normas universalmente aceptadas para muchas abreviaturas y acrónimos. En el caso del reconocimiento de abreviaturas, se mantiene una lista de palabras recortadas reconocidas [s46].

Pierre Vinken, 61 años, se unirá a la junta como director no ejecutivo 29 noviembre. El Sr. Vinken es presidente de Elsevier NV, el grupo editorial holandés.	<pre>... <Palabra inicio="0" fin="5" /> <Palabra inicio="7" fin="12" /> <Palabra inicio="15" fin="16" /> <Palabra inicio="20" fin="23" /> ...</pre>
---	---

Figura 2.3: Ejemplo de separación de palabras en un párrafo, obtenido de [s46].

En el ejemplo anterior, se obtiene la lista de palabras del texto, con la separación por palabras indicada por los espacios en blanco y los signos de puntuación.

3. Etiquetado gramatical o Part-of-Speech (POS) - tagging. Posterior a la ejecución de las dos técnicas previamente explicadas, se puede realizar el proceso de etiquetado de palabras de acuerdo con el rol que cumplen dentro de una oración. El proceso de Etiquetado gramatical consiste en asignar a cada una de las palabras de un texto su categoría gramatical de acuerdo con la definición de la misma o el contexto en el que aparece, como lo pueden ser los sustantivos, adjetivos, adverbios, etc.

Para lograr lo anterior, se necesita establecer las relaciones de una palabra con sus adyacentes dentro de una frase o de un párrafo. Un mismo token puede tener múltiples etiquetas POS, pero solo una es válida dependiendo del contexto.

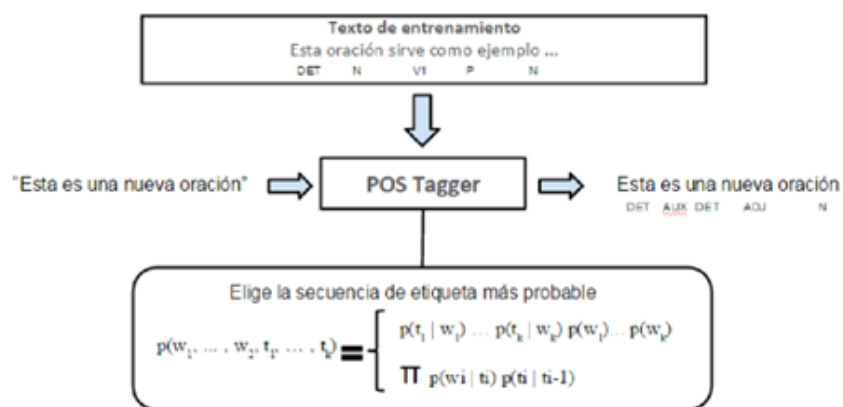


Figura 2.2. POSTagger.

Figura 2.4: POS Tagger, obtenido de [s46].

4. Segmentación morfológica. Un morfema se define como un fragmento mínimo capaz de expresar el significado de una palabra, es decir, es la unidad significativa más pequeña de un idioma.

Los morfemas se clasifican en 2 categorías. Los morfemas independientes admiten cierta libertad fonológica del lexema:

- Pronombres: cuíde-se, di-le, él, ella.
- Preposiciones: desde, a, con, de.
- Conjunciones: y, e, o, pero, aunque.
- Determinantes: él, ella, ese, un, una.

Por otro lado, los morfemas dependientes van unidos a otra unidad mínima dotada de significado, conocidos como monemas, para completar su significado. Los tipos de morfemas dependientes son:

- a) Derivativos: estos morfemas son facultativos, es decir, añaden matices al significado de los lexemas.
 - Prefijos
 - Sufijos
 - Interfijos
- b) Flexivos: estos morfemas son constitutivos, es decir, señalan relaciones gramaticales y sus accidentes entre los diferentes agentes de una acción verbal o una expresión nominal.
 - Género
 - Número
 - Persona
 - Modo y tiempo

La identificación de monemas permite el análisis en profundidad de una palabra en un texto, ya que de esta forma se obtiene información específica como el género, modo, tiempo, etc. y es posible ubicar de manera precisa cada palabra de una oración.

5. Eliminación de Stop Words. Mediante esta técnica, se excluyen palabras comunes que tienen poco valor para la recuperación de información. La cantidad de ocurrencias de una palabra en un texto determina si es o no una “stop word”, siendo que cuanto más ocurrencias existan menos relevancia tiene en el texto; en su mayoría, los artículos, los pronombres, las preposiciones y las conjunciones. La técnica de eliminación de stop words permite reducir el tamaño de un texto, eliminando aproximadamente el 30 % o 40 % de las palabras, y se seleccionan las palabras claves.

A partir de un listado de palabras Stop words, se detectan dentro del texto para su posible eliminación. En ocasiones, al listado de palabras de uso común se le agrega un conjunto de palabras propias del documento que se analiza, empleando la técnica TF-IDF (Term Frequency - Inverse Document Frequency), que permite determinar qué palabras son importantes para un documento de acuerdo con la frecuencia de aparición dentro de un texto.

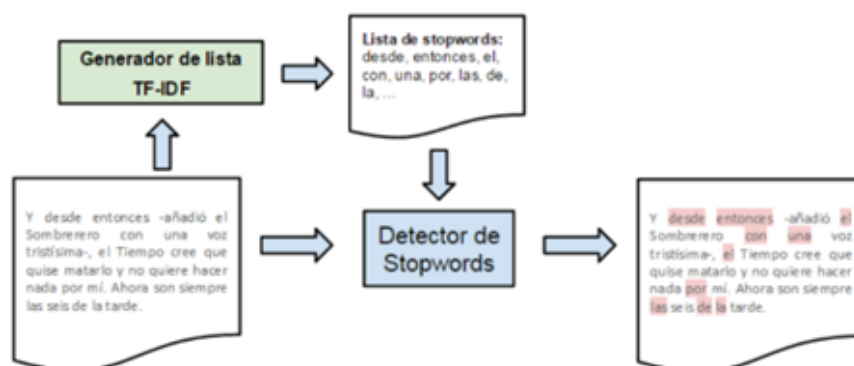


Figura 2.5: Ejemplo de Detección de Stop words, obtenido de [s34].

6. Reconocimiento de Entidades Nombradas (NER). Busca y clasifica elementos de texto que pertenecen a categorías predefinidas, como lo son nombres de personas, entidades, organizaciones, lugares, expresiones temporales, cantidades, porcentajes. etc. Para poder hacer el reconocimiento de las diferentes entidades, se utilizan una serie de aproximaciones, y para poder realizar el análisis es necesario tener una noción del contexto en el cual se encuentra cada una de las entidades para determinar a qué se refieren. Finalmente, dentro de las posibles entidades se realiza una asociación con los conceptos del contexto dentro de una base de datos de conocimiento.

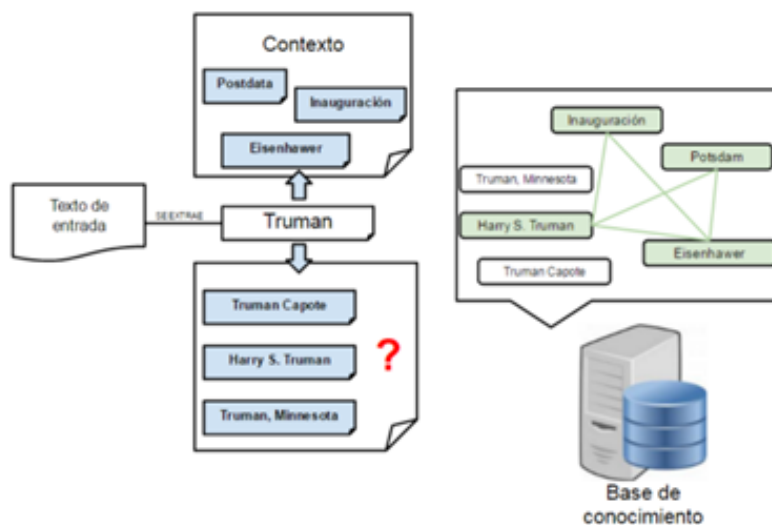


Figura 2.6: Reconocimiento de Entidades Nombradas, obtenido de [s46]

7. Stemming. Esta técnica busca un concepto de una palabra mediante la eliminación de prefijos y sufijos para obtener la raíz. De esta manera, se reduce la palabra a su mínimo elemento con significado

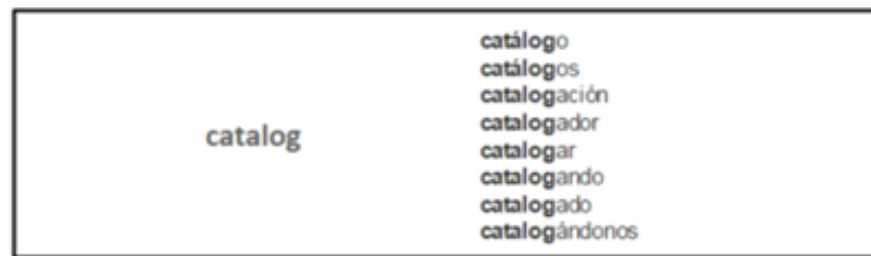


Figura 2.7: Ejemplo de los términos derivados de la raíz “catalog”, obtenido de [s46].

No obstante, es importante mencionar que esta técnica no siempre funciona correctamente debido a que hay palabras que poseen raíces compartidas por más de un significado.

Término con prefijo	Raíz/Stem	Término con el que causaría confusión
Prevalencia	valenc	Valencia, valencia, valenciano, ambivalencia, polivalencia
Precatalogar	catalog	Descatalogar, catálogo

Tabla 2.1: Ejemplos de términos con raíces compartidas, obtenido de [s46]

2.5.3. Aplicaciones del PLN

A continuación, se enlistan las principales aplicaciones del PLN:

- Recuperación y extracción de información. La recuperación de información (RI), es el proceso de encontrar en un repositorio grande de datos, material (usualmente documentos) de naturaleza no estructurada (usualmente texto) o semiestructurada (páginas Web, por ejemplo), que satisfaga una necesidad de información [s45].
- Por su parte, la extracción de información (EI) consiste en la obtención de las partes de las partes que interesan en el texto para pasarlas a un formato de base de datos [s45].
- Minería de datos. La minería de datos proporciona herramientas poderosas para descubrir patrones ocultos y relaciones en datos estructurados. Este proceso asume que los datos ya se encuentran almacenados en un formato estructurado. La minería de datos usa técnicas y metodologías de reconocimiento de información, extracción de información y corpus procesados con técnicas de lingüística computacional [s45].

- Sistemas de búsqueda de respuestas. Son sistemas diseñados para tomar una pregunta en lenguaje natural y proporcionar una respuesta, para evitar que los usuarios naveguen y lean varias páginas de resultados de búsqueda. Estos sistemas son construidos sobre motores de búsqueda y requieren contenido fuente para descubrir las respuesta; a su vez, deben tener métodos para entender las preguntas del usuario y determinar el tipo de respuesta que debe dar [s45].
- Generación de resúmenes automáticos. La generación de resúmenes consiste en emplear herramientas de PLN, dichos resúmenes pueden ser con enfoque extractivo o abstractivo. Por un lado, los métodos extractivos consisten en tomar una colección de términos, frases o párrafos significativos que definen el significado del texto original para generar un resumen. Por otro lado, los métodos abstractivos dependen de técnicas de parafraseo para producir las síntesis, las técnicas aún están siendo desarrolladas [s45].
- Análisis de sentimientos. El análisis de sentimientos en textos es la identificación y extracción de información subjetiva. Este proceso generalmente involucra el uso de herramientas de PLN y software de análisis de textos para automatizar el proceso. La forma básica de análisis de sentimientos es una clasificación polarizada de sentimientos que puede asignar calificaciones de en un rango de -10 a 10 que se basa en el aprendizaje para evaluar emociones tanto negativas como positivas en corpus etiquetados de entrenamiento [s45].
- Traducción automática. Consiste en tomar el texto escrito en un lenguaje y traducirlo a otro, manteniendo el mismo significado. El proceso de traducción automática sigue tres pasos: en primer lugar, el texto en el lenguaje original se transforma a una representación intermedia, luego se realizan modificaciones a esta representación intermedia basándose en la morfología del lenguaje, y finalmente se transforma al lenguaje destino [s45].

2.6. MediaPipe

MediaPipe es un conjunto de herramientas de código abierto para ser empleadas en tareas como el reconocimiento facial, seguimiento de gestos, detección de objetos y el seguimiento del cuerpo humano [s5].

2.6.1. Herramientas de MediaPipe

Las principales herramientas que ofrece MediaPipe son:

- MediaPipe Detección de caras. Permite detectar y seguir rostros de una imagen o vídeos en tiempo real, empleando técnicas de machine learning para mejorar la precisión [s5].

- Malla facial MediaPipe. Proporciona una malla 3D del rostro, para proporcionar información precisa sobre los rasgos faciales, lo cuál es útil en aplicaciones de animación y modelado 3D [s5].
- MediaPipe Seguimiento manual. Con esta herramienta se puede detectar y seguir los movimientos de la mano en tiempo real, con alta precisión [s5].
- MediaPipe Holistic. Combina la detección facial, el seguimiento de manos y el seguimiento corporal en una sola herramienta integrada, lo que es útil para aplicaciones de realidad aumentada y juegos [s5].
- MediaPipe Objectron. Es una herramienta para detectar y seguir objetos 3D en el espacio, siendo útil para comprender e interactuar con objetos reales en un entorno virtual [s5].
- Segmentación MediaPipe Selfie. Permite segmentar a las personas en el fondo de una imagen o vídeo [s5].
- MediaPipe Pose. Detecta las posturas del cuerpo humano, proporcionando información sobre las posiciones de las articulaciones y las extremidades [s5].
- Reconocimiento de gestos MediaPipe. Herramienta empleada en el reconocimiento de gestos de la mano para interacciones intuitivas y control de gestos [s5].
- MediaPipe EfficientDet. Mediante el uso de Redes Neuronales rápidas y eficaces, se puede mejorar la detección y localización de objetos en imágenes [s5].

2.6.2. **MediaPipe Pose**

Como se comentó en el apartado anterior, MediaPipe Pose permite detectar puntos de referencia de cuerpos humanos en una imagen o vídeo. Se emplea principalmente para identificar ubicaciones claves del cuerpo, analizar la postura y categorizar los movimientos [s51].

El marcador de poses emplea una serie de modelos para predecir los marcadores de poses [s51]:

- Modelo de detección de poses. Detectar la presencia de cuerpos con algunos puntos de referencia de poses clave.
- Modelo de marcador de pose. Agregar una asignación completa de una pose, en la que se generan 33 puntos de referencia de la pose 3D.

El modelo de marcador de pose realiza un seguimiento de 33 ubicaciones de puntos de referencia del cuerpo.

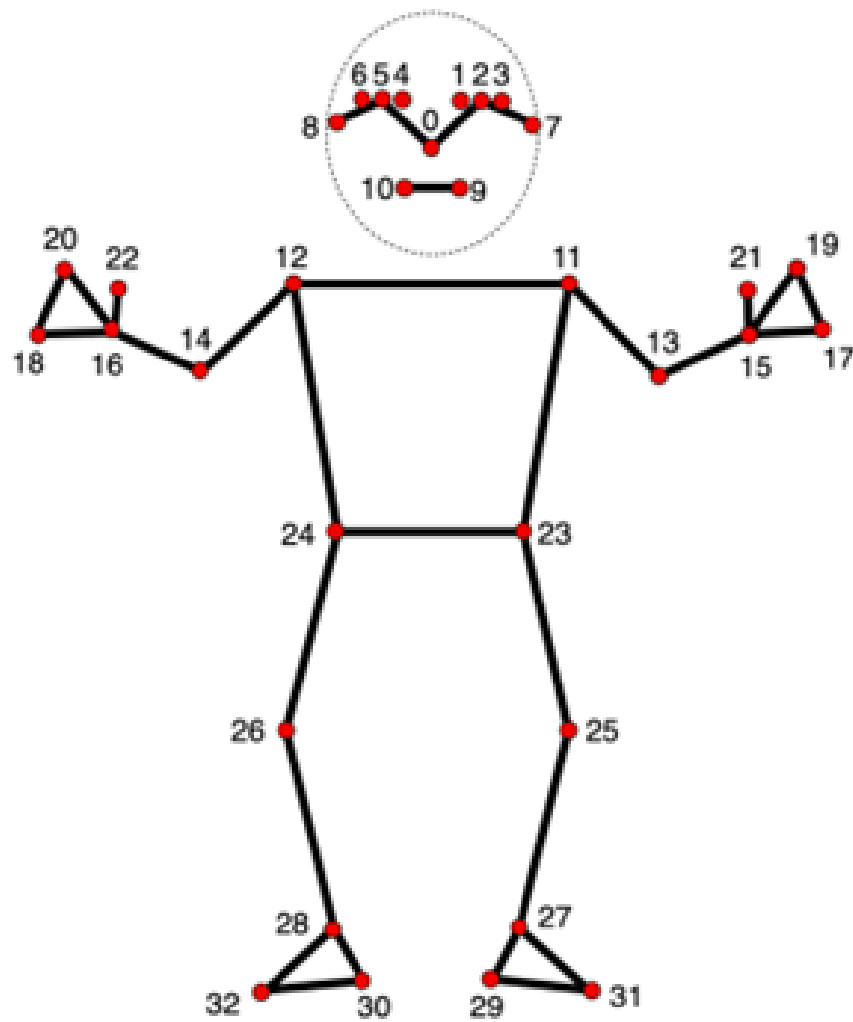


Figura 2.8: Ubicaciones de puntos de referencia del cuerpo [s51].

A continuación, se enlistan las partes representadas del cuerpo:

1. Nose - nariz
2. Left eye (inner) - ojo izquierdo (interior)
3. Left eye - ojo izquierdo
4. Left eye (outer) - ojo izquierdo (exterior)
5. Right eye (inner) - ojo derecho (interior)
6. Right eye - ojo derecho
7. Right eye (outer) - ojo derecho (exterior)

8. Left ear - oreja izquierda
9. Right ear - oreja derecha
10. Mouth (left) - boca (izquierda)
11. Mouth (right) - boca (derecha)
12. Left shoulder - hombro izquierdo
13. Right shoulder - hombro derecho
14. Left elbow - codo izquierdo
15. Right elbow - codo derecho
16. Left wrist - muñeca izquierda
17. Right wrist - muñeca derecha
18. Left pinky - meñique izquierdo
19. Right pinky - meñique derecho
20. Left index - índice izquierdo
21. Right index - índice derecho
22. Left thumb - pulgar izquierdo
23. Right thumb - pulgar derecho
24. Left hip - cadera izquierda
25. Right hip - cadera derecha
26. Left knee - rodilla izquierda
27. Right knee - rodilla derecha
28. Left ankle - tobillo izquierdo
29. Right ankle - tobillo derecho
30. Left heel - talón izquierdo
31. Right heel - talón derecho
32. Left foot index - punta del pie izquierdo
33. Right foot index - punta del pie derecho

2.7. Modelado de Animaciones 3D

El término animación 3D se refiere a la técnica de animación empleada para desplazar modelos tridimensionales generados digitalmente, sirviéndose para ello de un eje de coordenadas cartesiano virtual [s7].

La animación 3D ha estado históricamente más orientada a la replicación de la física del mundo real, ya que representa con total libertad la fuerza de gravedad, la inercia o la masa de cuerpos [s7].

2.7.1. Unity

Unity es una plataforma para el desarrollo de videojuegos y aplicaciones interactivas, que ofrece una amplia variedad de herramientas y recursos para crear experiencias visuales y funcionales [s8]. Es un motor gráfico empleado para desarrollar videojuegos, aplicaciones interactivas en 2D, 2D, realidad aumentada (AR) y realidad virtual (VR), y tiene soporte en varias plataformas como PC, consolas, dispositivos móviles y dispositivos de realidad extendida [s8].

Unity destaca por su conjunto de características robustas que facilitan el desarrollo de aplicaciones interactivas de alta calidad para la simulación física y el rendering, las cuáles requieren visualización y experiencia de usuario de alta calidad [s8].

En la actualidad Unity es empleado en múltiples industrias, además del desarrollo de videojuegos, ya que es popular en sectores como la arquitectura, el diseño automotriz, la medicina y la educación.

2.8. Sistema Operativo Android

Un sistema operativo móvil es un conjunto de programas que habilitan características específicas de un teléfono móvil y brindan servicios a las aplicaciones móviles que se ejecutan en él [s81].

El sistema operativo Android es un sistema operativo móvil desarrollado por la empresa estadounidense Google y que está basado en el sistema operativo Linux. Es un sistema operativo abierto, gratuito, versátil, seguro y altamente personalizable que está desarrollado para dispositivos móviles como smartphones y tablets [s81].

Las principales características de Android son las siguientes [s81]:

- Interfaz de Usuario (UI) personalizable. Los usuarios son capaces de cambiar el aspecto de sus dispositivos para adaptarlos a sus necesidades.

- Compatibilidad con múltiples fabricantes. Este sistema operativo es ejecutado en una gran cantidad de dispositivos de múltiples fabricantes.
- Google Play Store. Android cuenta con una tienda de aplicaciones que permite descargar diferentes aplicaciones de diversa índole, basadas en las necesidades de cada usuario.
- Asistente Virtual. Los usuarios tienen acceso a un asistente virtual llamado Google Assistant, que ayuda en la realización de tareas.
- Integración con servicios de Google. Android está integrado con servicios de Google como Gmail, Google Drive, Google Photos, Maps, entre otros.
- Compatibilidad con tecnologías emergentes. Android es compatible con tecnologías como la Realidad Virtual (VR), la Realidad Aumentada (AR) y los asistentes virtuales.

Capítulo 3

Desarrollo

3.1. Análisis y diseño de sistemas

3.1.1. Descripción del alcance

3.1.2. Entregables

3.1.3. Reglas de negocio

- **RN01:** El sistema debe permitir al usuario ingresar texto de forma manual o pegarlo desde otra aplicación.
- **RN02:** El sistema debe permitir solo la entrada de texto en español.
- **RN03:** El campo de entrada debe tener un límite de caracteres (por ejemplo, 30 caracteres).
- **RN04:** El sistema debe validar el texto ingresado para evitar caracteres no permitidos.
- **RN05:** El botón de traducción debe estar habilitado solo cuando el campo de texto no esté vacío.
- **RN06:** El sistema debe mostrar un mensaje de error si el texto ingresado no es válido.
- **RN07:** El sistema debe permitir al usuario editar el texto antes de enviarlo a traducción.
- **RN08:** El sistema debe traducir el texto ingresado a LSM.
- **RN09:** El sistema debe asignarle una única traducción y animación a cada texto validado.

- **RN10:** El sistema deberá deletrear el texto ingresado utilizando el alfabeto dactilológico de LSM si no existe una traducción en LSM en la base de datos.
- **RN11:** El sistema debe mostrar el resultado de la traducción en formato animado.

3.1.4. Requerimientos funcionales

- **RN01: Entrada de Texto** La aplicación debe permitir al usuario ingresar texto en español que será traducido a LSM.
- **RN02: Procesamiento de Lenguaje Natural (PLN)** Implementar un módulo que analice y procese el texto ingresado, identificando frases clave y contextos específicos para su traducción adecuada.
- **RN03: Generación de Avatares 3D** Desarrollar avatares 3D que representen visualmente las señas correspondientes a las frases procesadas. Los avatares deben ser capaces de mostrar expresiones faciales y movimientos corporales que reflejen la gramática y sintaxis de la LSM.
- **RN04: Animaciones y Transiciones** Las animaciones entre señas deben ser fluidas y naturales, utilizando IA para optimizar las transiciones y evitar movimientos bruscos o poco realistas.
- **RN05: Interfaz de Usuario (UI)** Diseñar una interfaz intuitiva y accesible que permita a los usuarios ingresar texto personalizado para su traducción.
- **RN06: Modo de Deletreo** En caso de que una frase no esté disponible en la base de datos, la aplicación debe ofrecer la opción de deletrear palabra por palabra utilizando el alfabeto dactilológico de la LSM.
- **RN07: Compatibilidad** La aplicación debe ser compatible con la última versión de Android, garantizando su funcionamiento en una amplia gama de dispositivos móviles.

3.1.5. Requerimientos no funcionales

- **RnF01: Rendimiento** La aplicación debe cargar y procesar las traducciones en menos de 2 segundos, garantizando una experiencia de usuario rápida y eficiente.
- **RnF02: Usabilidad** La interfaz debe ser fácil de navegar, con instrucciones claras y botones de acción bien definidos, siguiendo las heurísticas de usabilidad de Nielsen.

- **RnF03: Escalabilidad** El sistema debe ser capaz de adaptarse a futuras expansiones, como la inclusión de más frases, soporte para otros dialectos de señas o integración con otras plataformas.

3.1.6. Casos de uso

Caso de uso 01: Entrada de texto





Resumen

El usuario ingresa texto en español en la aplicación, que luego es procesado y traducido a LSM.

Descripción

Atributos	Tipo de texto: Español, límite de 500 caracteres, validación de caracteres permitidos.
Actores	Usuario (quien ingresa el texto).
Precondiciones	El sistema está abierto y funcionando correctamente. El usuario debe estar en la pantalla principal.
Actividad que se va a bitácora	Registro del texto ingresado por el usuario y la fecha y hora de la traducción solicitada.
Registro que se va a bitácora	Texto ingresado, usuario que lo ingresó, IP del dispositivo, timestamp de la acción.
Viene de	

Trayectoria principal

- 1  El sistema muestra un campo de entrada de texto en la pantalla principal de la aplicación, permitiendo al usuario ingresar texto en español. Además, la interfaz incluye:
 - Botón para iniciar la traducción.
- 2  Ingresa el texto en el campo de entrada.
- 3  El sistema procesa el texto ingresado y lo traduce a LSM, mostrando el resultado en formato animado o en video.
- 4  Puede modificar el texto y volver a ejecutar la traducción si es necesario.

— *Fin del caso de uso.*




Diseño de la interfaz**Caso de uso 02: Procesamiento de Lenguaje Natural (PLN)****Resumen**

El sistema procesa el texto ingresado mediante PLN y lo prepara para su conversión a señas en LSM.

Descripción

Atributos	Tipo de texto procesado, análisis semántico y sintáctico.
Actores	Sistema (encargado del procesamiento del texto).
Precondiciones	El texto ingresado por el usuario debe estar disponible y correctamente formateado.
Actividad que se va a bitácora	Registro del texto procesado, las frases clave y el análisis semántico realizado.
Registro que se va a bitácora	Texto procesado, identificaciones de frases clave, timestamp del análisis.
Viene de	El caso de uso . ^{Entrada de texto} .

Trayectoria principal

- 1  El sistema recibe el texto ingresado por el usuario.
- 2  El sistema analiza el texto utilizando el módulo de procesamiento de lenguaje natural, identificando frases clave y contexto semántico para una traducción adecuada.
- 3  El sistema prepara el texto procesado para la generación de las señas en LSM.

— *Fin del caso de uso.*

Caso de uso 03: Generación de Avatares 3D




Resumen

El sistema genera un avatar 3D que realiza y muestra en pantalla la traducción en señas LSM, incluyendo expresiones faciales y movimientos corporales.

Descripción

Atributos	Avatar 3D, animación de señas, expresiones faciales y movimientos corporales.
Actores	Sistema (generador del avatar 3D), Usuario (quien visualiza el avatar).
Precondiciones	El texto debe haber sido procesado y traducido a LSM correctamente.
Actividad que se va a bitácora	Registro de la generación del avatar y las animaciones realizadas.
Registro que se va a bitácora	Avatar generado, tipo de señas representadas, timestamp de la generación.
Viene de	El caso de uso "Procesamiento de Lenguaje Natural (PLN)".

Trayectoria principal

- 1  El sistema genera un avatar 3D correspondiente a la traducción de la frase en LSM.
- 2  El avatar 3D realiza las señas correspondientes, incluyendo expresiones faciales y movimientos corporales.
- 3  El avatar 3D se presenta en la pantalla al usuario, mostrando la traducción de la frase en LSM.

— *Fin del caso de uso.*




Caso de uso 04: Modo de Deletreo**Resumen**

El sistema detecta una palabra desconocida y activa el modo de deletreo, mostrando cada letra con gestos del alfabeto dactilológico de LSM.

Descripción

Atributos	Alfabeto dactilológico, letras individualmente representadas en LSM.
Actores	Usuario (quien observa el deletreo), Sistema (que activa el modo de deletreo).
Precondiciones	La palabra no debe estar en la base de datos de LSM.
Actividad que se va a bitácora	Registro del uso del modo de deletreo por parte del usuario.
Registro que se va a bitácora	Palabra deletreada, secuencia de letras, timestamp de la actividad.
Viene de	El caso de uso . ^{Entrada de texto.} "Procesamiento de Lenguaje Natural (PLN)".

Trayectoria principal

- 1  El sistema detecta que la palabra ingresada no está en la base de datos de LSM.
- 2  El sistema activa el modo de deletreo, permitiendo al usuario ver la palabra deletreada utilizando el alfabeto dactilológico de LSM.
- 3  El usuario observa las letras del deletreo en la pantalla, cada una representada por un gesto del alfabeto dactilológico de LSM.

— *Fin del caso de uso.*

Caso de uso 05: Animaciones y Transiciones




Resumen

El sistema usa IA para gestionar animaciones fluidas entre señas en LSM, evitando movimientos bruscos y mejorando la experiencia visual.

Descripción

Atributos	Animación fluida, transiciones entre señas optimizadas, uso de IA.
Actores	Sistema (que gestiona las animaciones y transiciones).
Precondiciones	El avatar 3D debe estar generado y las señas deben estar listas para la animación.
Actividad que se va a bitácora	Registro de la generación y ejecución de las animaciones y transiciones.
Registro que se va a bitácora	Transiciones entre señas, optimización de IA, timestamp de la animación.
Viene de	El caso de uso "Generación de Avatares 3D".

Trayectoria principal

- 1  El sistema gestiona las animaciones entre las señas de LSM, asegurando que sean fluidas y naturales.
- 2  Utiliza inteligencia artificial (IA) para optimizar las transiciones entre las señas, evitando movimientos bruscos o poco realistas.
- 3  El sistema presenta las animaciones de las señas con transiciones suaves, brindando una experiencia visual coherente.

— *Fin del caso de uso.*



Caso de uso 06: Compatibilidad**Resumen**

La aplicación es compatible con la última versión de Android y se adapta a diversos dispositivos y tamaños de pantalla.

Descripción

Atributos	Compatibilidad con versiones de Android, ajustes automáticos según dispositivos.
Actores	Sistema (que garantiza la compatibilidad), Usuario (que utiliza la aplicación).
Precondiciones	La aplicación debe estar instalada en un dispositivo con la versión compatible de Android.
Actividad que se va a bitácora	Registro de la versión de Android y la compatibilidad del dispositivo.
Registro que se va a bitácora	Versión de Android, tipo de dispositivo, timestamp de la verificación de compatibilidad.
Viene de	El caso de uso Inicio de la aplicación. ^o "Verificación de configuración del dispositivo".

Trayectoria principal

- 1  La aplicación debe ser compatible con la última versión de Android.
- 2  El sistema garantiza su funcionamiento en una amplia gama de dispositivos móviles, ajustándose a diferentes tamaños de pantalla y capacidades del dispositivo.

— *Fin del caso de uso.*

Emergencia (3 a 5 palabras)	Saludo (2 a 4 palabras)	Agradecimiento /Respuestas (2 a 4 palabras)
Ayuda, por favor. Llama a la policía. Necesito un médico. Estoy herido/a. ¿Dónde está el hospital? Es una emergencia. ¿Puedes ayudarme? Necesito asistencia urgente. ¿Dónde está la salida? ¡Fuego! ¡Fuego!	Hola, ¿cómo estás? Buenos días. Buenas tardes. Buenas noches. ¿Qué tal? Mucho gusto. Encantado/a de conocerte. ¿Cómo te llamas? Me llamo [nombre]. Hasta luego.	Gracias. Muchas gracias. Te lo agradezco. Que amable. Estoy muy agradecido/a. Gracias por tu ayuda. Cuando quieras. No hay problema. No pasa nada. Es un placer.

3.1.7. Diagrama de casos de uso**3.1.8. Mockups del sistema****3.1.9. Identificación de riesgos****3.1.10. Evaluación de riesgos****3.1.11. Mitigación de riesgos****3.2. Selección de los datos****3.2.1. Recolección de datos****3.3. Preparación de los datos**

Capítulo 4

Conclusiones

Capítulo 5

Referencias

Capítulo 6

Trabajo relacionado

Capítulo 7

Desarrollo

Capítulo 8

Experimentos

Capítulo 9

Conclusiones y trabajo futuro