

Aalto University
School of Science and Technology
Faculty of Information and Natural Sciences
Degree programme of Computer Science and Engineering
Espoo (Finland)
Director: Timo Honkela

Universidad de Zaragoza
Centro Politécnico Superior
Informática e Ingeniería de Sistemas
Ingeniería Informática Superior
Zaragoza (España)
Tutor: Manuel González Bedía

Jorge Ramón Letosa

Modelo de Comprensión de Frases en un Micromundo y Posible Aplicación al Campo de la Traducción Máquina

Proyecto Fin de Carrera

Modelo de Comprensión de Frases en un Micromundo y Posible Aplicación al Campo de la Traducción Máquina

RESUMEN

El trabajo realizado está dividido en dos fases principales. Primero se estudia la comprensibilidad de sentencias escritas y la representación en planos bidimensionales de las situaciones referidas por dichas sentencias. Posteriormente, el modelo desarrollado se utiliza para estudiar su adecuación en tareas de Traducción Máquina en las que se ha de tener en cuenta información contextual.

En la primera fase del proyecto se desarrolla un modelo que posibilita la representación de situaciones en un micromundo utilizando el mapa auto-organizativo SOM y el procesamiento de frases sencillas que describen dichas situaciones mediante el uso de una red neuronal recurrente simple (SRN). El mapa SOM es entrenado sobre un conjunto de situaciones ejemplo que conforman la experiencia del sistema y posibilitan la representación en el espacio de salida SOM del modelo cualquier situación posible. A continuación, un conjunto de frases es entrenado mediante la red SRN para posibilitar la transformación de frases a situaciones, resultando en su conjunto un modelo que recibe como valor de entrada una frase describiendo una determinada situación y muestra como salida, en el caso ideal, la representación de dicha situación sobre el mapa SOM.

La red SRN permite contar con información contextual basada en la información recibida en entradas previas, por lo que el modelo desarrollado resulta adecuado para tareas relacionadas con el procesamiento de lenguaje y, más concretamente, con la comprensión de frases e historias. Además, el espacio SOM de salida del modelo resulta ser independiente tanto del lenguaje utilizado en la entrada y su estructura proposicional como del número de situaciones que forman el conjunto de entrenamiento o experiencia, posibilitando la formación de un espacio sobre el cual poder realizar razonamientos lógicos básicos como si de un diagrama Venn se tratara.

En los experimentos realizados se muestra cómo el modelo no sólo es capaz de comprender frases a las que había sido entrenado, sino que también es capaz de generalizar, comprendiendo frases que no habían sido presentadas en el entrenamiento y que describen situaciones nuevas.

En la segunda fase del proyecto se diseña e implementa un sistema basado en el trabajo realizado en la primera fase con el que se pretende estudiar qué posibilidades ofrece el modelo siendo base o componente de un modelo que permita llevar a cabo tareas de Traducción Máquina (MT), prestando especial atención a aquellos casos donde la frase o palabra de entrada puede presentar diversas traducciones y esta ambigüedad ha de resolverse atendiendo al contexto de la situación.

El espacio de salida SOM se utiliza ahora como espacio único de representación conceptual compartido por dos microlenguajes que difieren en la manera en la que relacionan los símbolos léxicos con los conceptos, uno de ellos basado en el español y el otro en el inglés. Dicho espacio situacional compartido por ambos microlenguajes realiza la función de componente intermedio en la tarea de traducción a partir del cual se extrae información contextual y conceptual de la situación descrita por la frase de entrada, útil para producir como salida la nueva frase traducida.

De manera paralela al entrenamiento del espacio SOM, dos nuevos mapas SOM de mismo tamaño, uno para cada microlenguaje, son entrenados tomando como entrada la información conceptual presente en las situaciones del conjunto experiencia. Cada uno de estos mapas presenta la organización de las palabras del microlenguaje en base a su relación conceptual, de forma que aquellas palabras que poseen mayor similitud conceptual y/o son utilizadas para expresar situaciones similares aparecen dispuestas más próximas en estos mapas de vocabulario, que se utilizan en el proceso de producción de la frase traducida.

Este proyecto es una extensión de un trabajo publicado junto con el profesor Timo Honkela en la Conferencia Internacional de Redes Neuronales Artificiales (ICANN) de Septiembre de 2010: *Elementary Logical Reasoning in the SOM Output Space*.

Abreviaturas y Acrónimos

SOM	Self-Organizing Map
SRN	Simple Recurrent Network
NLP	Natural Language Processing
PLN	Procesamiento de Lenguaje Natural
MT	Machine Translation

Índice general

Abreviaturas y Acrónimos	III
1. Introducción	1
1.1. Estructura del Proyecto	4
2. Modelado del Conocimiento para PLN	5
2.1. Caso Monolingüe	5
2.1.1. Definición del Micromundo	5
2.1.2. Representación de Situaciones de Micromundo	7
2.1.3. Valores de Confianza	10
2.1.4. Comprensión de Frases	10
2.2. Caso Bilingüe	12
2.2.1. Mapa Situacional	12
2.2.2. Mapas de Palabras	15
2.2.3. Cómo Obtener la Traducción	20
3. Experimentos	28
3.1. Formación de Situaciones	28
3.2. Entrenamiento de Frases	29
3.3. Traducción de Frases	32
4. Conclusiones y discusión	37
4.1. Discusión	39
4.2. Trabajo Futuro	40

A. Métodos	44
B. Elementary Logical Reasoning in the SOM Output Space	54

Índice de cuadros

2.1. Diecisiete eventos básicos en el micromundo y su significado.	6
2.2. Gramática del microlenguaje.	11
2.3. 42 conceptos básicos en el micromundo y su significado.	13
2.4. Gramática inglesa.	15
2.5. Gramática española.	16
3.1. Conjunto de nuevas frases mencionando situaciones ya presentes en el conjunto de entrenamiento.	31
3.2. Conjunto de nuevas frases mencionando nuevas situaciones.	31

Índice de figuras

2.1. Razonamiento lógico básico al nivel de salida SOM.	8
2.2. Los planos de componentes de un mapa situacional.	9
2.3. Conjunción de los dos eventos básicos “lcHome” y “lcConcert”.	9
2.4. Los planos de componentes de un mapa situacional.	14
2.5. Posible traducción a aplicar para cada palabra en ambos microlenguajes.	17
2.6. Correspondencia usada entre conceptos y palabras y organización de los conceptos. Los elementos en negrita son conceptos de primer nivel, los cuales están directamente relacionados con una palabra a su derecha, coloreada en rojo para el idioma inglés y en verde para el conjunto español. El concepto “person” es el único en todo el conjunto que no pertenece al subconjunto de conceptos de primer nivel y está relacionado directamente con una palabra.	19
2.7. Mapas de vocabulario para ambos microlenguajes, inglés y español.	21
2.8. Máscaras de rol para el microlenguaje español.	24
2.9. Mapas de vocabulario incluyendo información sobre roles para ambos microlenguajes, inglés y español.	25
3.1. Situaciones ejemplo definidas a partir de la combinación de proposiciones básicas.	28
3.2. Valores de la conjunción de todas las combinaciones de eventos p y q , donde $p \neq q$	29
3.3. Resultados de acierto para cada palabra al traducir frases en inglés. El color rojo indica el total de apariciones en las frases de entrada, mientras el azul el número total de aciertos.	33

3.4.	Resultados de acierto para cada palabra al traducir frases en español. El color rojo indica el total de apariciones en las frases de entrada, mientras el azul el número total de aciertos.	34
3.5.	Resultados de acierto al traducir frases en inglés. El color azul indica el número total de aciertos, el color rojo el número de errores ligeros y el color amarillo son errores graves.	35
3.6.	Resultados de acierto al traducir frases en español. El color azul indica el número total de aciertos, el color rojo el número de errores ligeros y el color amarillo son errores graves.	36
A.1.	Arquitectura SOM.	45
A.2.	Celdas vecinas, mostrando un decrecimiento monótono. En el paso k , la función $N_q(k)$ cubre las celdas vecinas coloreadas en gris, naranja y rojo. Para el siguiente paso, $k + 1$, el conjunto de vecinos cubre las celdasnaranjas y rojas y, finalmente, para el paso $k + 2$ el conjunto ha sido reducido a las celdas rojas.	46
A.3.	Arquitectura SRN.	50
A.4.	Arquitectura SRN en este trabajo. La capa de entrada muestra una representación localista de cada palabra, mientras los vectores coloreados expresan representación distribuida.	53

Capítulo 1

Introducción

La mayor parte de las frases describen un escenario de acontecimientos o eventos en el mundo. De la comprensión de una de esas frases se deriva la construcción de una representación mental de ese escenario o situación a la cual se refiere la frase. Esta representación situacional es comparable a lo que Johnson-Laird [24] llama un *modelo mental*, y tiene la característica de que no es lingüística sino que está basada en la experiencia del agente que comprende y que tiene un determinado conocimiento sobre el mundo.

Aunque el objetivo del lector o la persona que escucha es construir una representación situacional, la mayoría de los modelos de comprensión de lenguaje se centran en las estructuras gramaticales o proposicionales del lenguaje, sin involucrar o involucrando mínimamente conocimiento sobre el mundo.

En este trabajo, se presenta un modelo que no tiene en cuenta las estructuras proposicionales. En su lugar, el modelo implementado transforma cada frase de entrada en su correspondiente representación de la situación.

El proyecto está dividido en dos estudios principales, el segundo de los cuales viene motivado por los resultados obtenidos en el primero. En la primera de las partes, una red neuronal recurrente simple (SRN) es utilizada para llevar a cabo comprensión de frases sencillas, entendidas como secuencia de palabras, construidas a partir de la definición de una microgramática en inglés que permite construir un conjunto de 396 posibles frases, cada una de las cuales describe una situación. Estas representaciones de una situación se obtienen mediante el entrenamiento no supervisado de un conjunto de situaciones ejemplo proposicionalmente definidas sobre un 'self-organizing map' (SOM) de tamaño fijo. Los principales objetivos de este primer estudio son:

1. Cómo representar conocimiento sobre situaciones mediante el uso de mapas de Kohonen (SOM).
2. Cómo procesar y representar frases mediante el uso de una red SRN.
3. Cómo evaluar la comprensibilidad de frases entrenadas así como analizar los resultados en situaciones de entrada de nuevas frases y nuevas combinaciones de palabras que no habían sido entrenadas.
4. Mostrar cómo se puede llevar a cabo razonamiento lógico elemental usando el espacio de salida del SOM.

Los resultados obtenidos en los experimentos muestran cómo el modelo satisface los dos tipos de generalización que se le exigen a un sistema de simulación de comprensión de frases. Además de ser capaz de comprender aquellas frases a las que ha sido entrenado previamente, es capaz también de comprender frases que no le han sido presentadas anteriormente. Por un lado, el sistema comprende nuevas descripciones de las situaciones a las que ha sido entrenado. Por ejemplo, si el modelo ha aprendido a construir la representación situacional para la frase “BD and LC write a book”, también es capaz de generar la misma representación para la frase “LC and BD write a book”. Por otro lado, el modelo es capaz de comprender frases que describen situaciones a las que no ha sido entrenado.

Teniendo este primer modelo como partida y tras analizar los resultados de los experimentos se pretende vincular este primer modelo con el estudio de “Machine Translation”. Un sistema de *traducción inteligente* de frases debería llevar a cabo diversas tareas más o menos complejas que vayan más allá de la simple traducción individual de las palabras que componen la frase de entrada. El proceso de traducción comienza por la *comprensión de la frase* a traducir para, posteriormente, producir una nueva frase expresada en un idioma diferente. Esta comprensión implica conocer cuál es el significado de la frase como un todo, por lo que un modelo de traducción debe ser capaz de comprender la función que cada palabra tiene en una determinada frase y cómo el contexto puede hacer variar su significado. Existen ya modelos de traducción que son capaces con mayor o menor precisión de traducir frases y utilizar la información contextual presente para producir la nueva frase traducida, pero en la gran mayoría de los casos son modelos que procesan la frase a nivel léxico o gramatical, sin tener en cuenta la idea de *comprensión*. Sin embargo, la contextualización de una frase o palabra no es el único problema que surge en la traducción. El lenguaje de una persona está basado en

su propia experiencia subjetiva como individuo. Por ejemplo, es más probable que una persona no experta describa un fenómeno en términos generales, mientras un experto utiliza términos más específicos. De esta forma, incluso aunque dos personas hablen en mismo idioma no se puede asumir que el vocabulario de ambos sea exactamente el mismo. La correspondencia existente entre los símbolos léxicos o palabras y el mapa conceptual del mundo para un agente es diferente en cada caso.

Esta afirmación se encuentra presente con mayor importancia todavía cuando dos idiomas diferentes son considerados. La tarea de traducción ya no sólo ha de resolver adecuadamente la ambigüedad de una frase o palabra gracias al contexto, sino que ha de ser capaz también de encontrar la correspondencia más precisa en el mapa conceptual del lenguaje al que se quiere traducir.

Un nuevo modelo basado en el primer estudio es diseñado en el cual dos nuevas microgramáticas, una en inglés y otra en español son definidas. Un único SOM es ahora entrenado a partir de situaciones ejemplo expresadas en cualquiera de los dos microlenguajes dando lugar a un único espacio común de representación de situaciones. La tarea de traducción consiste básicamente en transformar una frase expresada en un determinado lenguaje (inglés o español) en su correspondiente traducción al otro lenguaje utilizando el espacio SOM como elemento intermedio del cual se extrae información relacionada con el contexto y los elementos que se encuentran presentes en una determinada representación situacional que se obtiene como salida a la entrada de una frase. Los objetivos de este segundo estudio incluyen:

1. Cómo crear mediante un SOM un espacio único de razonamiento para dos lenguajes cuyos modelos conceptuales son distintos mediante el uso de conceptos simples.
2. Analizar la organización conceptual que aparece en un SOM de un conjunto de palabras que forman un microlenguaje.
3. Llevar a cabo y evaluar la traducción de frases entre dos lenguajes y en ambas direcciones, prestando especial interés a aquellos casos en los que la palabra a traducir presenta varias posibles traducciones dentro del segundo idioma y esta ambigüedad ha de resolverse atendiendo al contexto en el que ésta se utiliza.

1.1. Estructura del Proyecto

Tras la introducción del capítulo 1, en el capítulo 2 se explican los métodos seguidos así como los componentes y modelos desarrollados y utilizados. El capítulo 3 se centra en los experimentos llevados a cabo para evaluar el sistema y los resultados obtenidos. Por último, en el capítulo 4 se pueden encontrar las conclusiones del proyecto.

El Anexo A ofrece información relativa a los dos métodos principales usados durante el proyecto, los mapas auto-organizativos SOM y las redes neuronales recurrentes SRN. Además de incluir la base teórica de dichos métodos, se ofrece una visión de cómo éstos han sido útiles en trabajos relacionados en el campo de “Machine Translation”, Procesamiento del Lenguaje Natural y representación del conocimiento. En el Anexo B se incluye una copia del artículo publicado junto con el profesor Timo Honkela [16] que resume parte del trabajo realizado durante la primera fase del proyecto.

Capítulo 2

Modelado del Conocimiento para Procesado de Lenguaje Natural

2.1. Caso Monolingüe

Para implementar conocimiento sobre el mundo, comenzaremos construyendo un marco definido para un micromundo en el que un conjunto de situaciones tiene lugar. Después de definir el micromundo y sus limitaciones, se explicará cómo el conocimiento sobre un micromundo puede ser representado por medio de la ayuda de los mapas auto-organizativos (Self-Organizing Maps).

2.1.1. Definición del Micromundo

Desde el momento en el que consideramos a un lector como algo más que un simple decodificador de signos y tenemos en cuenta el rol activo de comprender, la importancia del conocimiento previo del lector aparece. Para construir artificialmente este conocimiento inicial sobre el mundo, se sigue el esquema propuesto por Frank [3], de forma que se lista un conjunto de eventos básicos (Fig. 2.1) del cual cualquier situación contenida dentro del marco del micromundo puede ser obtenida.

En este micromundo hay dos protagonistas diferentes “LC” y “BD” que pueden realizar tres acciones diferentes: dar un concierto, escribir un libro, y grabar un CD. Además, pueden estar en su país natal y/o en Europa, pueden tener éxito o no como resultado de sus acciones y pueden estar acompañados durante sus actividades en el micromundo. De esta manera, combinando uno o más de estos eventos al mismo tiempo que se satisfacen las restricciones

Cuadro 2.1: Diecisiete eventos básicos en el micromundo y su significado.

No.	Name	Meaning
1	70s	In the 70s decade.
2	80s	In the 80s decade.
3	lcSuccess	LC gets success.
4	bdSuccess	BD gets success.
5	lcFails	LC fails.
6	bdFails	BD fails.
7	lcWrites	LC writes a book.
8	bdWrites	BD writes a book.
9	lcRecords	LC records a CD.
10	bdRecords	BD records a CD.
11	lcConcert	LC performs a concert.
12	bdConcert	BD performs a concert.
13	lcHome	LC is at the home country.
14	lcEurope	LC is in Europe.
15	bdHome	BD is at the home country.
16	bdEurope	BD is in Europe.
17	accompanied	He is/they are accompanied.

del micromundo, es posible crear las proposiciones necesarias que permiten representar el conocimiento del mundo y las situaciones (como será explicado posteriormente). A modo de ejemplo, combinando los eventos número uno ("70s"), cuatro ("bdSuccess") y ocho ("bdWrites"), construimos una proposición en la cual BD escribe con éxito un libro en los setenta.

Siguiendo el modelo presentado en [5], el conocimiento del micromundo no está implementado explícitamente sino a partir de un entrenamiento sobre un mapa auto-organizativo sobre un conjunto de 227 situaciones ejemplo (cada una de ellas creadas a partir de la combinación de eventos como ha sido ejemplificado anteriormente). De esta forma, el modelo refleja de manera las regularidades inherentes, limitaciones y probabilidades *a priori* de ocurrencia de los eventos concretos del micromundo. Esto sirve como un tipo de "experiencia" que el sistema tiene después de ser entrenado con un conjunto de situaciones en el micromundo. Como ejemplo, algunas de las restricciones del modelo son que como mucho una acción puede ser desempeñada en una situación, que los protagonistas del modelo nunca escriben un libro juntos o que cada vez que "LC" escribe un libro se encuentra en su país natal. También hay algunas otras regularidades en el modelo que reflejan la probabilidad de ocurrencia de ciertas situaciones: es más probable que "BD" logre éxito en la década de los ochenta que en la década de los setenta.

2.1.2. Representación de Situaciones de Micromundo

Una proposición p puede ser representada en un área bidimensional de manera que es posible obtener la probabilidad de ocurrencia de dicha situación y de cualquier combinación de situaciones p y q por medio de teoría de conjuntos. La figura 2.1 muestra un ejemplo donde p y q son dos proposiciones con una probabilidad de 0.5 cada una en el escenario de un micromundo. Para cada proposición, un área que ocupa la mitad del área total del rectángulo refleja la probabilidad *a priori* de 0.5 de ocurrir para cada proposición. Teniendo en cuenta las siguientes ecuaciones, cualquier combinación lógica de proposiciones en el escenario del micromundo puede ser representada.

$$Pr(\neg p) = 1 - Pr(p). \quad (2.1)$$

$$Pr(p \wedge q) = Pr(p)Pr(q). \quad (2.2)$$

De hecho,

$$Pr(p \vee q) = Pr(p) + Pr(q) - Pr(p \wedge q), \text{ y } Pr(p \oplus q) = Pr(p) + Pr(q) - 2Pr(p \wedge q).$$

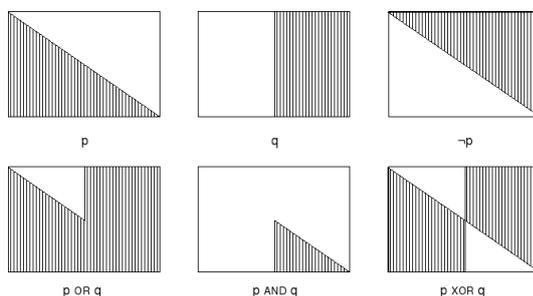


Figura 2.1: Razonamiento lógico básico al nivel de salida SOM.

Siguiendo esta idea, un Self-Organizing Map (SOM, [14]) es usado para construir automáticamente la representación del mapa de las situaciones del micromundo. Una situación en el micromundo está representada por un vector de 150 componentes. Esta implementación en la que una situación está contenida en un vector de tamaño fijo de n elementos permite una representación en la que n es independiente del tamaño del conjunto de situaciones ejemplo, aunque, al mismo tiempo, al reducir la dimensión desde k , tamaño del conjunto de situaciones ejemplo, a $n = 150$, parte de la información contenida en las situaciones ejemplo se puede perder. Un vector de situación está en la forma $s(p) = (s_1(p), s_2(p), \dots, s_n(p))$, donde cada $s_i(p)$ tiene un valor entre 0 y 1 para cada evento p (p puede ser, de hecho, cualquier combinación de eventos $s_i(p)$) que indica la medida para la cual cada componente, o más precisamente, la celda en el SOM, es parte de la representación de p (véase [3]). La figura 2.2 muestra el mapa representacional para cada proposición básica del micromundo. Como se puede observar en la figura, el patrón que representa *lcHome* se solapa con el patrón de *lcWrites*, mostrando que si LC está escribiendo un libro, LC está en su país natal.

En este punto, se advierte que el área de una proposición en una representación SOM no está estrictamente definida sino de manera difusa. Por esta razón la teoría de conjunto difusos tiene que ser usada ahora para definir las correspondientes áreas de las diferentes combinaciones lógicas de proposiciones, de forma que el valor de cada celda en el mapa es dado por:

$$s_i(\neg p) = 1 - s_i(p). \quad (2.3)$$

$$s_i(p \wedge q) = s_i(p)s_i(q). \quad (2.4)$$

De las ecuaciones 2.3 y 2.4, es decir, negación y conjunción, todas las posibles combinaciones lógicas pueden ser obtenidas. De hecho, la disjunción p ó q

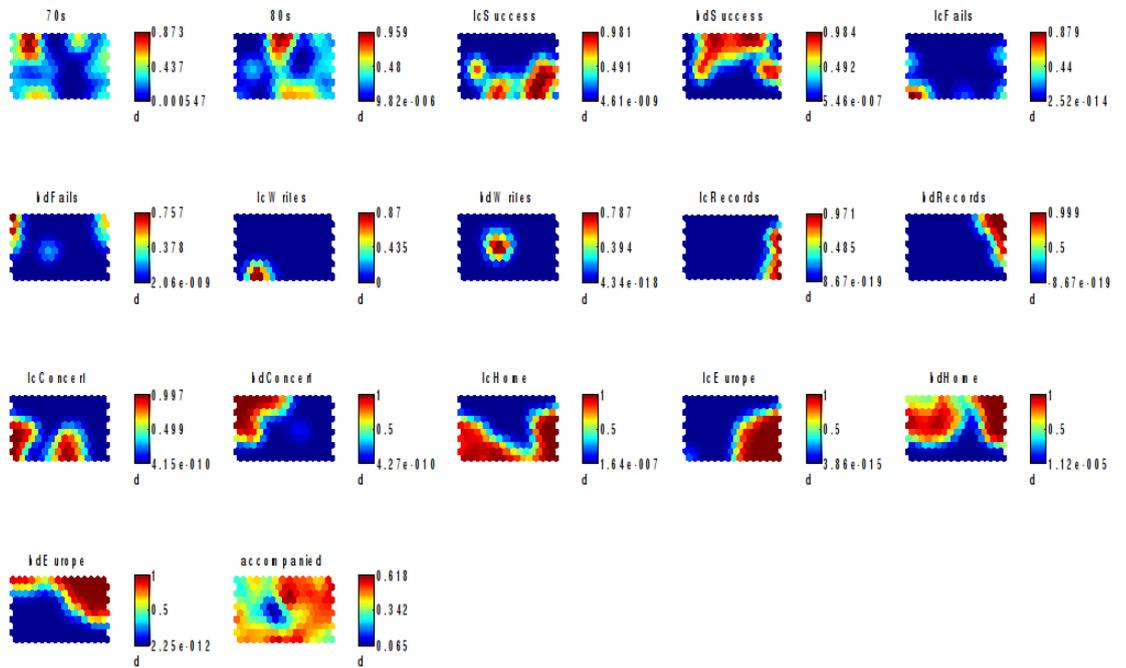


Figura 2.2: Los planos de componentes de un mapa situacional.

está definida por $s_i(p \vee q) = s_i(p) + s_i(q) - s_i(p)s_i(q)$, mientras $s_i(p \oplus q) = s_i(p) + s_i(q) - 2s_i(p \wedge q)$.

La figura 2.3 muestra una situación ejemplo en la que LC da un concierto en su país natal. Esta situación es el resultado de la conjunción (ecuación 2.4) de situaciones “lcConcert” y “lcHome”.

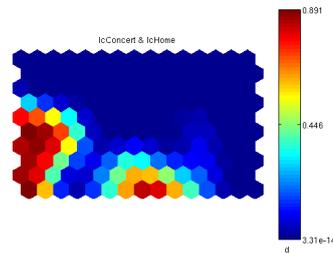


Figura 2.3: Conjunción de los dos eventos básicos “lcHome” y “lcConcert”.

2.1.3. Valores de Confianza

Ya que no existe correspondencia uno a uno entre proposiciones y dimensiones en la representación tomada de situaciones, usamos *belief values* para calcular los resultados dados como vectores de situación. Definimos $X = (x_1, x_2, \dots, x_n)$ como un vector de situación, con n el número de dimensiones del espacio de situaciones. Como resultado de entrenar el SOM, la probabilidad (la estimación de la probabilidad es aproximada) de que la situación X suceda en el micromundo es igual a la fracción del mapa que cubre (véase [3]). El valor de confianza τ de una situación X está definido del siguiente modo:

$$\tau(X) = \frac{1}{n} \sum_i x_i. \quad (2.5)$$

Además, del hecho de que $Pr(p|X) = Pr(p \wedge X)/Pr(X)$, se puede obtener la probabilidad subjetiva de una determinada proposición p dado que la situación X es el caso.

El valor de confianza de p en la situación X se define como:

$$\tau(p|X) = \frac{\sum_i s_i(p)x_i}{\sum_i x_i}. \quad (2.6)$$

2.1.4. Comprensión de Frases

Llegados a este punto se requiere una manera de comprender un pequeño conjunto de frases simples. En otras palabras, es necesario convertir frases de un microlenguaje a la representación de la correspondiente situación en el micromundo.

Microlenguaje

El microlenguaje está formado por las siguientes 13 palabras: *LC, BD, and, is, gets_success, fails, performs_concert, writes_book, records_CD, at_home, in_Europe, single, accompanied*. Con estas palabras y siguiendo las reglas de la gramática en la figura 2.2, 396 frases diferentes pueden ser obtenidas refiriéndose, cada una de ellas, a una situación del micromundo.

Cuadro 2.2: Gramática del microlenguaje.

S	→	NP VP
NP	→	LC BD LC and BD BD and LC
VP	→	Action [Place and is State and Result]
	→	is Place [and Action and State and Result]
	→	is State [and Action Place and Result]
	→	Result [and Action Place and is State]
Action	→	writes_book performs_concert records_CD
Place	→	at_home in_Europe
		at_home and in_Europe in_Europe and at_home
State	→	single accompanied
Result	→	gets_success fails

Desde las Frases hasta las Situaciones

Las frases del microlenguaje son transformadas en representaciones situacionales en forma de vector por medio del entrenamiento de una red neuronal recurrente simple [2]. La capa de entrada de la red está formada por 13 unidades, una por cada palabra en el microlenguaje. La capa oculta tiene 50 unidades y la capa de salida 150 unidades, una por cada dimensión del espacio de situaciones.

Las palabras de una frase son procesadas una por una, de manera que sólo una de las unidades de entrada está activa en cada momento. De esta forma, la salida de la red se puede observar en cada paso y es posible analizar cómo la representación de las situaciones se define a medida que la frase de entrada se completa.

Resultados de la Comprensión

Después de entrenar la red, los valores de confianza que fueron explicados anteriormente resultan útiles para medir el rendimiento del modelo. La idea es que la situación $X(P)$ representada por la salida de una determinada frase deberían tener un valor de confianza mayor que el valor de confianza *a priori* $\tau(p)$ de la situación p correspondiente a esa frase:

Resultado de comprensión for p :

$$compr(p) = \frac{\tau(p|X(p)) - \tau(p)}{\tau(p|p) - \tau(p)}. \quad (2.7)$$

Cuando $\tau(p|X(p)) > \tau(p)$, el resultado de comprensión es mayor que 0 y

la red refleja que la frase ha sido comprendida, siendo el caso ideal cuando $\tau(p|X(p)) = \tau(p|p)$, y el resultado es 1. Por otra parte, cuando el resultado es negativo la red malinterpreta la frase, y no hay comprensión cuando el resultado es igual a cero.

2.2. Caso Bilingüe

2.2.1. Mapa Situacional

La idea es similar a la ya implementada en el caso monolingüe [16]. Un conjunto de situaciones ejemplo conforma los datos de entrada de un único mapa SOM del que se espera que muestre la representación de las situaciones. Sin embargo, en este caso, cada una de las situaciones del conjunto de situaciones ejemplo no está representada exactamente como una combinación de eventos básicos en el micromundo. La idea de evento básico es ahora cambiada a la idea de *concepto básico* de forma que una situación ejemplo se construye por la combinación de conceptos o características de dicha situación. Algunos ejemplos de *conceptos básicos* pueden incluir: persona, instrumento, juego, resultado, etc... La lista completa de 42 *conceptos básicos* usados en el caso bilingüe en este trabajo aparece en la tabla 2.3.

En este punto, el proceso es idéntico al ya llevado a cabo en la sección 2.1.2 del caso monolingüe para obtener las representaciones situacionales. En este caso, un conjunto de 433 situaciones ejemplo, cada una de ellas creada de combinar los *conceptos básicos* como ha sido ya visto para el caso monolingüe es la entrada para un único mapa auto-organizativo de 432 unidades. Los mapas resultantes para cada uno de los *conceptos básicos* es mostrado en la figura 2.4.

Los mismos razonamientos lógicos presentados para el caso monolingüe son ahora aplicables a las representaciones situacionales de la figura 2.4. Además, los resultados también muestran la relación que aparece entre los diferentes *conceptos básicos* de la tabla 2.3. De hecho, es posible ver cómo las celdas más activas de los conceptos “piano” (14) y “guitar” (17) están incluidas en la región que el concepto “instr” (15) cubre, en concordancia con la definición del micromundo en la cual las ideas “piano” y “guitar” son consideradas un instrumento.

Cuadro 2.3: 42 conceptos básicos en el micromundo y su significado.

No.	Name	Meaning	No.	Name	Meaning
1	person	Person.	22	have	To have.
2	man	Man.	23	fail	To fail (concepts 24 and 25).
3	woman	Woman.	24	miss	To miss a game or fail a performance.
4	playgame	To play a game/sport.	25	failexam	To fail a test.
5	playinstr	To play a musical instrument.	26	test	Test.
6	earn	To earn something.	27	takeout	To take out.
7	win	To win a game/competition.	28	success	Success.
8	bank	Bank.	29	exam	Exam.
9	bench	Bench (to sit on it).	30	pass	To pass.
10	football	Football.	31	chair	Chair (to sit on it).
11	sport	Sport.	32	rest	To rest.
12	game	Game.	33	letter	Letter (written message).
13	cards	Playing cards.	34	gift	Gift (present).
14	piano	Piano.	35	getrec	To receive.
15	instr	Instrument.	36	seat	Seat.
16	music	Music.	37	present	Present (gift).
17	guitar	Guitar.	38	result	Result of an action.
18	money	Money.	39	building	Building.
19	sit	To sit.	40	banking	The banking sector.
20	get	To get something.	41	mail	Mail.
21	obtain	To obtain something.	42	ownership	Ownership.

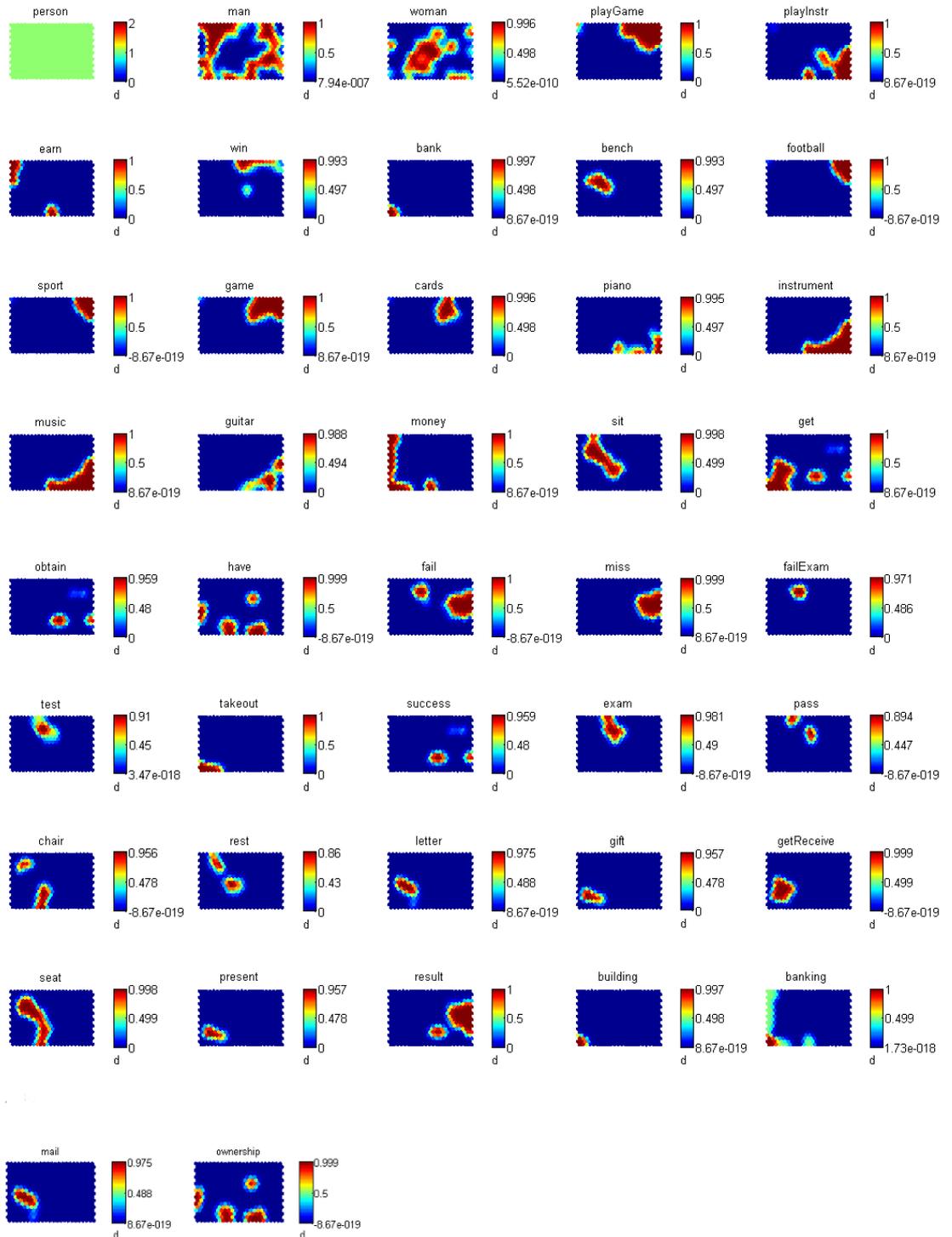


Figura 2.4: Los planos de componentes de un mapa situacional.

2.2.2. Mapas de Palabras

El lenguaje de una persona está basado en las experiencias subjetivas del individuo [10]. De hecho, es más probable que una persona no experta describa un evento en términos generales, mientras un experto usa términos más específicos [10]. De esta manera, el mapa conceptual de una persona y los correspondientes enlaces con el conjunto de símbolos de su vocabulario es subjetivo y, por consiguiente, puede ser diferente para cada individuo que comparta el mismo lenguaje.

Este asunto desemboca en un estudio más difícil y desafiante del procesamiento del lenguaje natural cuando varios lenguajes diferentes son considerados en la comunicación.

Microlenguaje

Los dos lenguajes diferentes usados en este estudio son el español y el inglés. De forma similar a lo que sucedía en el caso monolingüe, se necesita una microgramática para construir las frases referidas a su correspondiente situación en el micromundo. De esta forma, cada posible situación en el micromundo puede ser expresada en cualquiera de los dos lenguajes diferentes por medio del uso de su microgramática. La definición de la gramática inglesa es mostrada en la tabla 2.4, mientras la española se puede ver en 2.5.

Cuadro 2.4: Gramática inglesa.

S	→ NP VP
NP	→ Joe Sophie Somebody
VP	→ plays [Game] [Result]
	→ plays Instr [Result2]
	→ Result passes rests
	→ gets [Object]
	→ sits [Seat] [Rest]
	→ has Sth
	→ does_exam [ResExam]
Game	→ football cards
Result	→ wins fails gets success earns money
Instr	→ piano guitar
Result2	→ fails gets success earns money
Object	→ gift letter money [Bank]
Bank	→ bank
Seat	→ chair bench
Rest	→ rests
Sth	→ piano guitar money chair letter gift cards
ResExam	→ passes fails

Cuadro 2.5: Gramática española.

S	→	NP VP
NP	→	Jose Sofía Alguien
VP	→	juega [Game] [ResultG]
	→	toca Instr [Result2]
	→	saca Object1
	→	recibe Object2
	→	se_sienta [Seat] [Rest]
	→	tiene Sth
	→	hace_examen [ResExam]
	→	resultG aprueba suspende descansa
Game	→	fútbol cartas
ResultG	→	gana fracasa tiene éxito gana dinero
Instr	→	piano guitarra
Result2	→	fracasa tiene éxito gana dinero
Object1	→	dinero [Bank]
Object2	→	regalo carta
Bank	→	banco
Seat	→	silla banco
Rest	→	descansa
Sth	→	piano guitarra dinero silla carta regalo cartas
ResExam	→	aprueba suspende

El microlenguaje inglés está formado por 24 palabras: *Joe, Sophie, somebody, has, plays, sits, does_exam, football, gets, piano, cards, guitar, letter, gift, money, chair, bench, wins, fails, passes, earns, rests, bank, success*. Por otra parte, para conseguir las frases equivalentes de la gramática inglesa en español, se necesitan 25 palabras: *Jose, Sofía, alguien, tiene, juega, se_sienta, hace_examen, fútbol, recibe, saca, guitarra, toca, carta, regalo, dinero, silla, banco, gana, fracasa, suspende, aprueba, descansa, éxito, piano, cartas*.

Los microlenguajes previos han sido diseñados de forma que algunas características interesantes puedan ser analizadas. La mayoría de las palabras en ambos microlenguajes tienen una única traducción directa en el otro microlenguaje; de hecho, palabras en el subconjunto inglés como *somebody, gift* y *chair* están enlazadas unívocamente con *alguien, regalo* y *silla* en el subconjunto español, respectivamente. Sin embargo, existen algunas otras palabras en ambos microlenguajes que conducen a una traducción ambigua que puede ser resuelta correctamente sólomente teniendo en cuenta el contexto de la situación en el que la palabra está siendo usada. Como ejemplo, el verbo inglés *to play* debería ser traducido en español como *jugar* cuando se refiere a un juego, un deporte o un rol mientras *tocar* tiene que ser usado cuando se refiere a un instrumento musical. La lista completa de palabras en ambos microlenguajes y la posible traducción que puede ser aplicada se muestra en

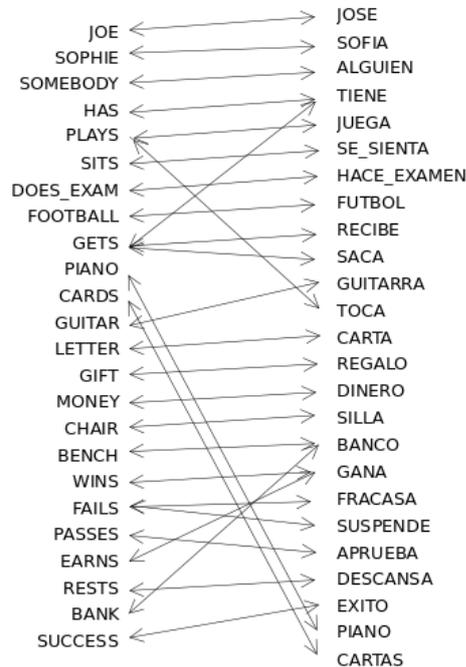


Figura 2.5: Posible traducción a aplicar para cada palabra en ambos microlenguajes.

la figura 2.5.

Apendizaje de Vocabulario

Para cada situación del conjunto de entrenamiento de entrada previamente descrito en la sección "Situational Map", una frase refiriéndose a esa situación es añadida. De esta forma, la entrada del modelo puede ser vista como un par *situation-sentence* que permite: (1) conseguir como salida un único mapa situacional basado en los *conceptos básicos* que componen cada situación y (2) relacionar las palabras que aparecen en cada frase con la situación emparejada así como con los *conceptos básicos* de la definición del micromundo.

Las situaciones ejemplo de entrada son vectores de la forma $S_i = (s_1, s_2, \dots, s_n)$, donde n es el número de *conceptos básicos*, $n = 42$, y $i = 1..m$, $m = 433$, el número de situaciones ejemplo en el conjunto de entrada de entrenamiento. De esta manera, en cada situación S_i , el evento $s_j (j = 1..n)$ es igual a 1 si el concepto j está presente en la situación i o 0 si no lo está. Hasta este punto ya es posible crear el mapa situacional mostrado en la figura 2.4.

Además, cada situación S_i está acompañada por una frase relacionada T_i

expresada en español o en inglés de forma que pueden existir situaciones repetidas expresadas en ambos lenguajes ($S_i = S_j, i, j = 1..n$ donde T_i es inglés (español) y T_j es español (inglés)) pero también situaciones expresadas sólo en uno de los microlenguajes ($S_i \neq S_j \forall j = 1..i-1, j = i+1..n$, o $S_i = S_j$ y T_i, T_j son ambas inglés (español)), y es también posible que ciertas clases de situaciones (situaciones diferentes con un subconjunto común de *conceptos básicos*) aparezcan con mayor frecuencia expresados en uno de los microlenguajes en lugar del otro. Es importante remarcar que el espacio de representación de situaciones se forma por la combinación de *conceptos básicos*, es decir, es independiente del lenguaje.

Una frase se define como la concatenación de palabras $W_i = (w_1, w_2, \dots, w_l)^T$, donde l es el número de palabras de cada microlenguaje, $l = 24$ para el caso inglés y $l = 25$ para el español. Se usa una representación localista de manera que cada palabra W_i en un microlenguaje está relacionada unívocamente con una posición en el vector, teniendo sólo una de las componentes marcadas como 1 en cada palabra y el resto a 0. De esta manera, una frase tiene la forma $T = (W_1, W_2, \dots, W_k)$, donde $k = 1..5$ es el número de palabras de la frase T .

La relación entre los elementos de cada par $[S_i, T_i]$ no es arbitraria sino que las palabras de una frase han de ser “asimiladas” por el sistema como los símbolos de los *conceptos básicos*. Así, la aparición de cierto $s_j \in S_i$ implica la existencia de cierto $W_k \in T_i$, y viceversa. Estos enlaces están definidos y se muestran en la figura 2.6. El conjunto de 42 elementos de *conceptos básicos* puede ser visto como un conjunto de árboles en los que pueden aparecer hasta tres niveles de abstracción. La existencia en una situación de un concepto de primer nivel implica la existencia de los conceptos de niveles superiores en ese árbol. Por ejemplo, cuando el concepto “man” es parte de una situación S_i , “person” también es parte, y además, la palabra “Joe” o “José” son parte de T_i , dependiendo del lenguaje de T_i .

Sin embargo, es importante aclarar que esta relación de la figura 2.6 es solamente útil a la hora de crear el conjunto de entrada del sistema, mientras lo que es visible para el sistema es sólo un vector de componentes $[S_i, T_i]$ sin información sobre los enlaces entre los elementos que constituyen cada par i .

En este punto, el sistema es capaz de relacionar una situación del micromundo con una frase expresada en alguno de los dos lenguajes pero lo que es necesario es conseguir averiguar el significado de una palabra.

Para alcanzar este objetivo se sigue el siguiente algoritmo:

Entonces, dos mapas diferentes se obtienen, cada uno de ellos teniendo como entrada el conjunto de palabras de cada microlenguaje. Cada vez que una palabra aparece relacionada con una situación, cada *concepto básico* parte de ella se añade al vector de 42 componentes de esa palabra. Este método implica que no existe ningún tipo de diferencia entre las palabras de una frase en el momento de asignar sus características, cada palabra recibe toda la información disponible contenida en la situación relacionada.

El resultado es que cada palabra W_j cuenta ahora con un conjunto de 42 valores, cada uno de ellos relacionado con un *concepto básico*, que mide la relación existente entre una palabra y cada uno de los *conceptos básicos*. Un valor de estos 42 es mayor para una palabra W_j si el *concepto básico* relacionado aparece con frecuencia como parte de S_i , donde $W_j \in T_i$ para el par de entrada $[S_i, T_i]$. Por otra parte, un valor es bajo para W_j si el correspondiente concepto no aparece con frecuencia como parte de S_i where $W_j \in T_i$ para el par de entrada $[S_i, T_i]$. Por ejemplo, para la palabra *guitar*, el concepto básico está altamente relacionado, pero los conceptos 'playInstr' o 'have' están relacionados también. Los mapas resultantes para cada microlenguaje se muestran en la figura 2.7.

Los mapas de la figura 2.7 presentan la relación de palabras para ambos microlenguajes después de seguir el algoritmo previo.

Esta idea está relacionada con el trabajo presentado por Honkela [10] en el que palabras de dos lenguajes diferentes son dispuestos en el mismo mapa auto-organizativo. En este trabajo, sin embargo, los valores de entrada de cada palabra no son asignados por medio de grados de relación subjetivos entre cada palabra y sus correspondientes valores conceptuales, sino que esta relación surge automáticamente del algoritmo.

2.2.3. Cómo Obtener la Traducción

La acción de traducir requiere llevar a cabo un proceso en el que se necesitan muchas subtareas y habilidades y no hay un único camino válido para hacerlo. De manera simple, la traducción de una frase comienza con la comprensión de la frase, que es diferente de la comprensión del significado de cada palabra individualmente, y la producción de una nueva frase en un lenguaje diferente. Existen muchas publicaciones y trabajos sobre comprensión de sentencias pero no es fácil encontrar estudios publicados sobre producción de frases.

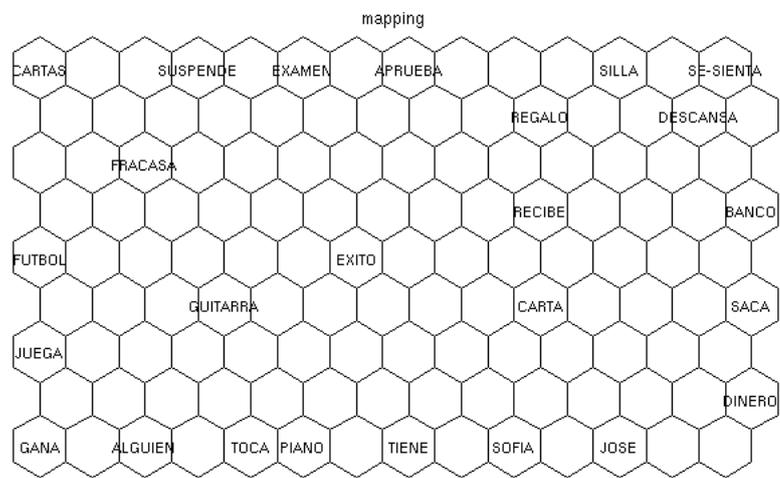
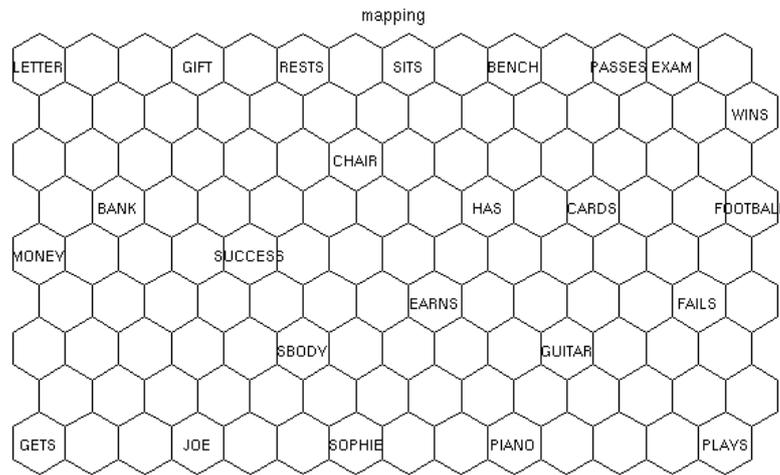


Figura 2.7: Mapas de vocabulario para ambos microlenguajes, inglés y español.

Roles en una Frase

Aquí, se define una estructura simple común de roles para producir la frase traducida. Hasta cinco roles pueden aparecer en una frase: *subject*, *action*, *object1*, *result* y *object2*. Esta estructura simple evita el uso de preposiciones, artículos y adverbios.

Las palabras de un microlenguaje pueden aparecer siempre actuando como un mismo rol o no. De hecho, la palabra *money* puede aparecer como 'object1' en una frase, pero también como 'object2' en otra.

Como ha sido explicado, el aprendizaje de vocabulario y los correspondientes resultados mostrados en la figura 2.7 vienen determinados por los *conceptos básicos* y su frecuencia de ocurrencia a lo largo del conjunto de entrenamiento. Esta clase de disposición es interesante [10] pero presenta dificultades a la hora de extraer automáticamente los elementos léxicos adecuados del mapa y usarlos en la tarea de producción de frases. La producción de frases requiere la correcta construcción de una secuencia de palabras en un lenguaje siguiendo las reglas de su gramática específica de modo que la intención comunicativa preverbal del hablante sea transformada en una secuencia ordenada de símbolos.

Ya que el objetivo principal de este trabajo es el estudio de traducción de palabras en casos 'directos' y en casos ambiguos y no la producción de frases, que es un problema extremadamente exigente, el modelo simple previamente mostrado de cinco roles es usado como gramática para ambos microlenguajes.

En este punto del estudio, los mapas de vocabulario carecen de información sobre las reglas de la gramática. Como respuesta a esta necesidad, una nueva variación del algoritmo es aplicada de forma que el vector de entrenamiento de entrada para cada palabra W_j está compuesto ahora por 47 variables, 42 componentes correspondiendo a *conceptos básicos* y 5 nuevas componentes que corresponden a cada uno de los cinco posibles roles:

```

for every pair  $[S_i, T_i]$  of the input set
  for every word  $W_j \in T_i$ 
     $R = (0, 0, 0, 0, 0)$ 
     $r_i = Rol(W_j, T_i)$ 
     $R(r_i) = C$ 
    if language = 'english' then
      index:= findWord( $W_j$ , EnglishVocabulary)
      inputEnglishMap'(index):= inputEnglishMap'(index) +  $S_i|R$ 
    else
      index:= findWord( $W_j$ , SpanishVocabulary)
      inputSpanishMap'(index):= inputSpanishMap'(index) +  $S_i|R$ 
    end if
  end loop
end loop
engMapRoles:= train(inputEnglishMap');
spaMapRoles:= train(inputSpanishMap');

```

inputEnglishMap' y **inputSpanishMap'** son vectores de 47 componentes, formados por la concatenación de un vector situación S_i y R . Cuanto mayor sea la constante C utilizada en el algoritmo, mayor importancia tendrán las variables relativas a los roles sobre la organización final, en detrimento de las variables conceptuales.

La figura 2.8 muestra la salida obtenida para las variables relativas a los roles en el lenguaje español. Dichas salidas son útiles posteriormente a modo de máscaras de rol en el proceso de producción de frases en la traducción.

La modificación descrita resulta en una nueva disposición de las palabras en cada mapa (ver figura 2.9) donde los elementos léxicos están ahora organizados en un primer nivel por roles y, en un segundo nivel, por los *conceptos básicos*.

El Proceso de Traducción

Llegados a este punto, todas las herramientas y elementos requeridos para el proceso de traducción han sido obtenidos. En esta sección se explica el método seguido y el uso de los componentes del modelo para conseguir la correcta traducción de una frase en cualquiera de los dos microlenguajes.

Es importante recordar una vez más que la intención de esta sección no es centrarse en un único sistema de traducción sino desarrollar un estudio de manera que las características y componentes que el modelo ofrece puedan ser

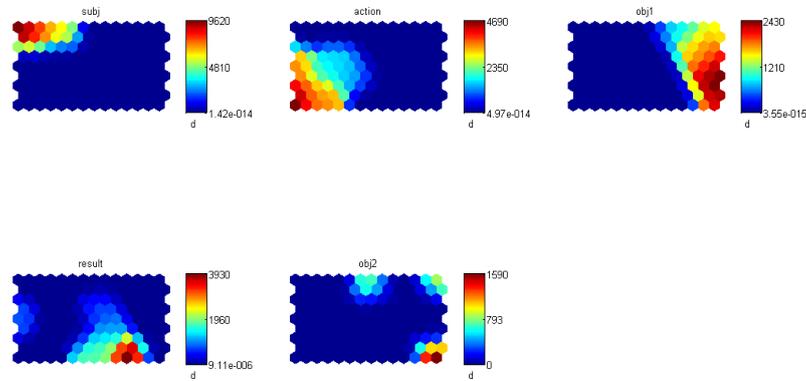


Figura 2.8: Máscaras de rol para el microlenguaje español.

analizadas, incluso aunque diferentes caminos de decisión hacia una solución concreta puedan quedar abiertos por ello.

El proceso seguido para traducir una frase es el siguiente:

1. Obtener la representación de la situación.
2. Extraer los componentes básicos del mapa de situación.
3. Obtener los valores correctos del mapa de palabras.
4. Aplicar las máscaras para cada rol.

1.- Obtener la representación de la situación. De manera similar a lo que sucedía en el caso monolingüe, cada frase, ya sea en inglés o en español, está relacionada con una representación situacional, que es un mapa auto-organizativo obtenido a partir de las correspondientes combinaciones lógicas.

2.- Extraer los componentes básicos del mapa de situación. El paso anterior crea una representación independiente del lenguaje a partir de una frase expresada en uno de los dos microlenguajes. La salida SOM obtenida es una matriz en la que cada componente (o celda del mapa μ_i) está etiquetada de forma que contiene información sobre si una situación del conjunto ejemplo de entrada es parte de ella o no. Además, cada componente posee un valor, $\mu_i \in [0, 1]$ indicando la medida en la que μ_i es parte de la representación situacional. Así, la celda con valor máximo de la matriz obtenida

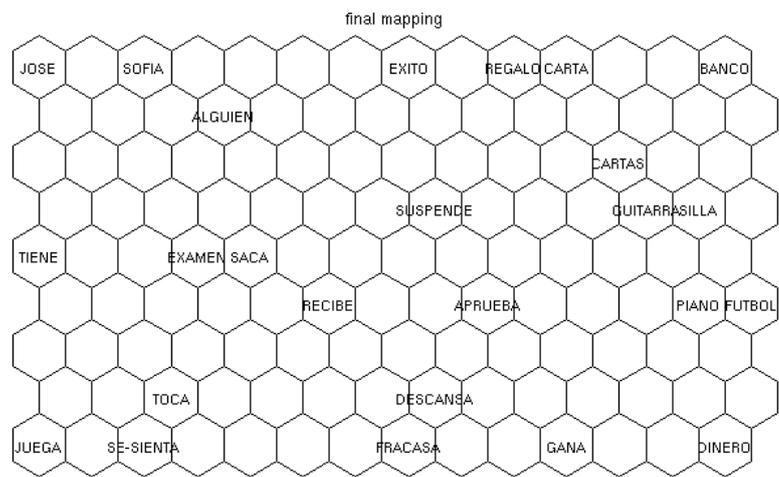
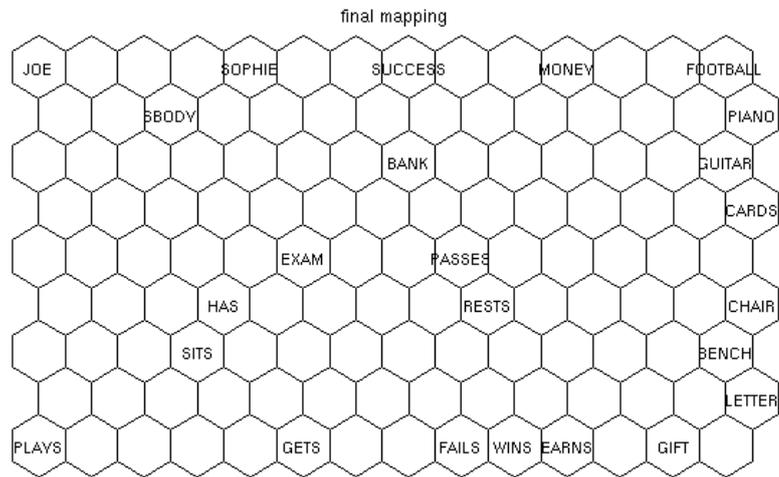


Figura 2.9: Mapas de vocabulario incluyendo información sobre roles para ambos microlenguajes, inglés y español.

contiene información sobre aquellas situaciones ejemplo del conjunto de entrenamiento que poseen una mayor relación con la situación a la que la frase de entrada hace referencia. Siguiendo esta idea, el algoritmo busca la celda con valor máximo que esté etiquetada y, entonces, se busca alguna etiqueta de una situación ejemplo que tenga tantos de conceptos básicos de primer nivel como número de roles tenga la frase de entrada. El algoritmo de búsqueda procede del siguiente modo:

1. La celda contiene alguna situación con el mismo número de conceptos básicos que roles hay en la frase. Si la celda contiene información sobre un único patrón de situación con el número de conceptos básicos requeridos, el algoritmo devuelve dicho patrón. Si más de un patrón diferente es encontrado, se devuelve el patrón más repetido. En el caso de que se encuentren diferentes patrones con la longitud requerida un mismo número de veces, uno de esos patrones es seleccionado aleatoriamente.
2. La celda contiene alguna situación con más conceptos básicos que roles. Si, al aplicar el caso previo, no se encuentra resultado, este segundo caso es aplicado. La misma forma de proceder que en el caso anterior es aplicada, ahora sobre patrones con más componentes de los que la frase de entrada requiere. Esto significa que se busca un patrón de situación más específico de lo que la frase de entrada requiere.
3. La celda contiene algún patrón de situación con menos conceptos básicos que roles. Si ninguno de los pasos previos ha encontrado resultado se aplica este caso con la misma metodología de búsqueda, ahora sobre patrones con menos componentes de los que requiere la frase de entrada. De esta manera, se procede a buscar un patrón de situación más general de lo que la frase de entrada requiere.

3.- Obtener los valores correctos del mapa de palabras. Se selecciona cada uno de los mapas correspondientes a los conceptos básicos obtenidos en el paso anterior para el microlenguaje objetivo. Entonces, se aplica la conjunción sobre el subconjunto de mapas de vocabulario. Cada mapa es, en realidad, un vector donde cada componente $x_i \in [0, 1]$, de forma que la conjunción se obtiene como resultado de la multiplicación escalar de todos los mapas seleccionados. Si los mapas de vocabulario de la figura 2.7 fueran usados, el vector resultante contendría información acerca de las palabras con más posibilidades para la traducción codificadas como aquellas componentes con los valores más altos; sin embargo, la tarea de construcción de una frase

correcta quedaría incompleta.

4.- Aplicar las máscaras para cada rol. De nuevo, diferentes caminos pueden seguidos en este punto para alcanzar el objetivo de la traducción. En esta solución, los roles presentes en la frase a ser traducida son tenidos en cuenta para conseguir la traducción correspondiente de cada rol. El siguiente algoritmo es usado:

```
translatedSentence:=  $\emptyset$ 
for every role  $r_i \in S$ 
    aux:= ApplyMask( $r_i$ , Solution)
    w:= GetWord(aux);
    Concatenate(translatedSentence, w)
end loop
```

Capítulo 3

Experimentos

En las siguientes líneas se describen los experimentos realizados en este trabajo incluyendo tanto la formación del mapa auto-organizativo de las situaciones como el modelo de red recurrente simple para las frases que describen situaciones en el micromundo.

3.1. Formación de Situaciones

El conjunto de situaciones ejemplo constituye la entrada del SOM y consiste en 227 situaciones ejemplo que siguen las limitaciones del micromundo de manera que cada situación en este conjunto se obtiene de una o más proposiciones combinadas. Estos son algunos ejemplos de posibles situaciones ejemplo correctas en nuestro micromundo:

$$\begin{aligned} &LC(writes) \wedge decade(70) \\ &BD(records) \wedge decade(80) \wedge BD(home) \wedge BD(success) \\ &LC(europe) \wedge BD(europe) \wedge accompanied \end{aligned}$$

Figura 3.1: Situaciones ejemplo definidas a partir de la combinación de proposiciones básicas.

Una situación de entrada es un vector $X = (x_1, x_2, \dots, x_v)$, $v = 17$, número de eventos en el micromundo. Una componente x_i es igual a 1 si es el caso de la situación ejemplo o 0 si no lo es. Esta representación implica, como se puede observar de la tabla 2.1, que el evento *accompanied* o su negación estarán

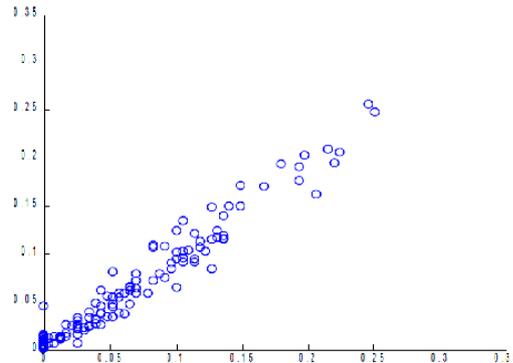


Figura 3.2: Valores de la conjunción de todas las combinaciones de eventos p y q , donde $p \neq q$.

siempre presentes en una situación ejemplo, mientras otras, como el caso de *getting success* o no (*failing*) permiten representar situaciones donde la información sobre el resultado de alguna acción por parte de los protagonistas del micromundo no es dada al modelo, es decir, *getting success* and *failing* tienen ambos un valor igual a 0. Después de entrenar el SOM con el conjunto ejemplo, un error de cuantificación final de 0.873 y un error topográfico final de 0.0 son obtenidos, mostrando el resultado de la figura 2.2.

Para comprobar la relación entre la probabilidad real inherente en el conjunto ejemplo con los valores de confianza de cada vector de situación en la salida, los valores de conjunción de cada evento p y q , son calculados en cada caso, mostrando los resultados de la figura 3.2. El coeficiente de correlación obtenido es muy alto, $r = 0,9773$, y no hay datos distantes del resto.

3.2. Entrenamiento de Frases

La red neural recurrente simple es entrenada con un conjunto de 368 frases que son presentadas a la red de manera aleatoria. El resto de frases (28) no son entrenadas y son usadas para medir el rendimiento de la red con frases que no fueron presentadas en el entrenamiento. Una frase se construye como la concatenación de las palabras que la componen, de manera que el resultado es una matriz de 13 filas, una por cada posible palabra en el microlenguaje, y c columnas, c el número de palabras contenidas en la frase. A modo de ejemplo, la ecuación 3.1 muestra la matriz correspondiente para la frase "*LC is single at_home*".

$$[lc, is, single, at_home] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.1)$$

El conjunto mencionado de 368 frases es presentado a la red recurrente durante 400 iteraciones de tal manera que el error cuadrático medio es calculado y retropropagado para encontrar los gradientes del error para cada peso y bias.

Después de entrenar y calcular los resultados de comprensión para las frases presentes en el conjunto de entrenamiento, el porcentaje de frases comprendidas es 89,7 y la media del resultado es 0,3931. Por otro lado, un conjunto de 20 frases no presentadas a la red durante el entrenamiento pero refiriéndose a situaciones en el micromundo que estaban ya presentes en el conjunto de entrenamiento son también consideradas, obteniendo una media de 0,3330 sobre las 18 de las 20 frases que fueron comprendidas. Estas 20 frases, mostradas en la figura 3.1, son frases no presentadas en el conjunto de entrenamiento pero que se refieren a situaciones ya referidas por otras frases presentes en el conjunto de entrenamiento. Por ejemplo, la frase "*LC is in_Europe and performs_concert*" no es parte del conjunto (segunda frase), mientras la frase "*LC performs_concert in_Europe.*" se encuentra en el conjunto de entrenamiento.

Finalmente, un conjunto de 8 frases no presentadas describiendo nuevas situaciones (figura 3.2) resultó en una media de 0,3799 y 7 del conjunto de 8 frases tuvieron un resultado mayor que 0. Estas ocho frases que no están presentes en el entrenamiento describen situaciones no descritas por ninguna otra frase en el conjunto de entrenamiento. Además, los resultados muestran cómo las frases cortas conducen hacia mayores resultados de comprensión que las largas.

Cuadro 3.1: Conjunto de nuevas frases mencionando situaciones ya presentes en el conjunto de entrenamiento.

No.	sentence
1	LC performs _concert in _Europe and a _home.
2	LC is in _Europe and performs _concert.
3	LC is accompanied and performs _concert.
4	LC fails and records _CD.
5	LC and BD records _CD at _home and in _Europe.
6	LC and BD is at _home and in _Europe.
7	LC and BD is in _Europe and gets _success.
8	BD is at _home and performs _concert.
9	BD is in _Europe and records _CD.
10	BD and LC performs _concert at _home and in _Europe.
11	BD and LC performs _concert in _Europe.
12	BD and LC performs _concert at _home.
13	BD and LC performs _concert and is single.
14	BD and LC performs _concert and is accompanied.
15	BD and LC is at _home and in _Europe.
16	BD and LC fails and records _CD.
17	BD and LC fails at _home and in _Europe.
18	BD and LC fails at _home.
19	BD and LC fails in _Europe.
20	BD and LC fails in _Europe and at _home.

Cuadro 3.2: Conjunto de nuevas frases mencionando nuevas situaciones.

No.	sentence
1	LC is accompanied and performs _concert.
2	LC performs _concert and is accompanied.
3	LC and BD is accompanied and performs _concert.
4	LC and BD performs _concert and is accompanied.
5	BD and LC performs _concert and is accompanied.
6	BD and LC is accompanied and performs _concert.
7	BD performs _concert and is accompanied.
8	BD is accompanied and performs _concert.

3.3. Traducción de Frases

Para analizar el rendimiento del sistema en el proceso de traducción de frases se presenta como entrada y se analizan los resultados obtenidos después de presentar todas las frases posibles creadas a partir de la gramática de cada microlenguaje.

Durante el proceso de ejecución de la tarea de traducción, la presencia de una frase en la entrada del modelo es transformada en la salida como la conjunción de las representaciones básicas de aquellos conceptos de la figura 2.4 que están ligados a las palabras de la frase de entrada, tal y como se muestra en la figura 2.6. Teniendo R como la representación que se obtiene en la salida a partir de una determinada frase de entrada, formada como la concatenación de palabras w_i , $w_1w_2..w_n : R = c_1 \cdot c_2 \cdot .. \cdot c_n$, donde c_i es la representación que se obtiene como salida para cada w_i . Además, $w_i : c_i = c_{i1} + c_{i2} + .. + c_{im}$, donde cada c_{ij} es la representación situacional que se obtiene para cada uno de los conceptos básicos diferentes ligados a una misma palabra.

Para analizar los resultados del experimento se comprueba, para cada frase de entrada T_i y cada $w_j \in T_i$, si la palabra obtenida como traducción de $w_j \in T_i$ es la traducción correcta y si ocupa el rol que le corresponde en la frase que resulta de la traducción. En caso de que así sea y ambas condiciones se cumplan, la traducción de w_i ha sido un acierto, en caso contrario ha sido un error.

En la figura 3.3 se muestran los resultados obtenidos al presentar todas las posibles frases del microlenguaje inglés. Los resultados muestran un acierto global del 82,61 %, porcentaje de acierto que se incrementa hasta el 88,91 % si no se tiene en cuenta la palabra *somebody* que, como ya se observó en la figura 2.6, es el único caso especial de palabra que no está relacionada con ningún concepto de primer nivel.

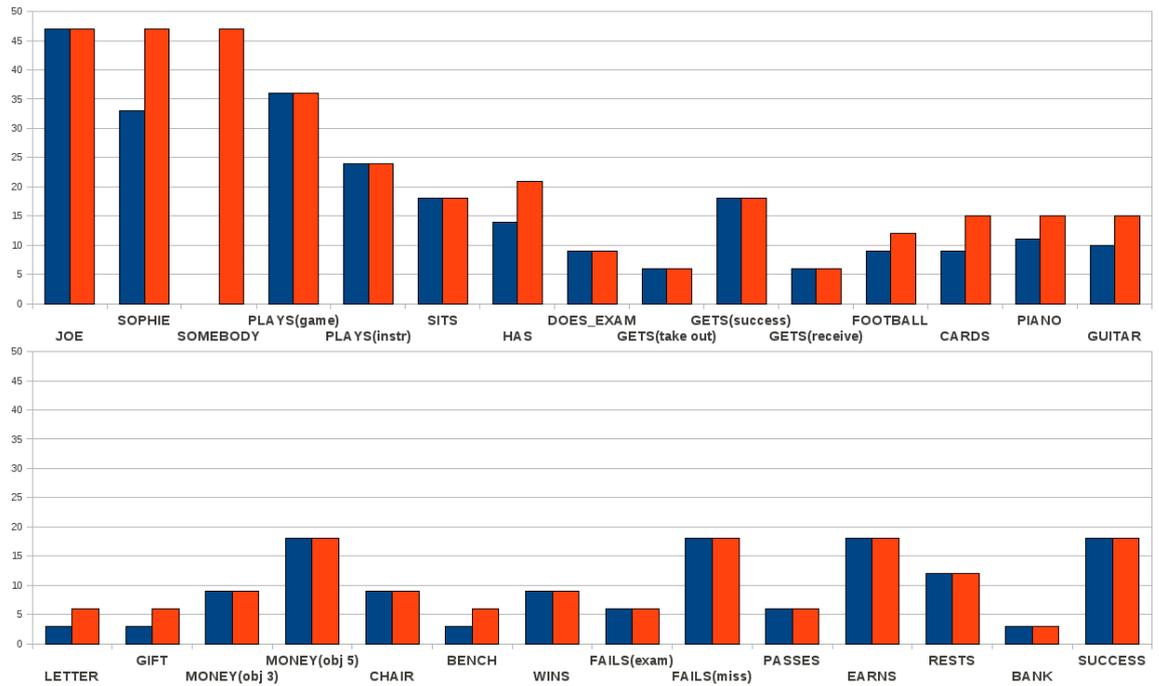


Figura 3.3: Resultados de acierto para cada palabra al traducir frases en inglés. El color rojo indica el total de apariciones en las frases de entrada, mientras el azul el número total de aciertos.

Por otra parte, al realizar el mismo experimento con la totalidad de las frases posibles creadas a partir de la microgramática española, se obtienen los resultados de la figura 3.4. Esta vez, se obtiene un porcentaje de acierto igual al 87,5%, que se ve incrementado hasta el 91,41% en el caso de no tener en cuenta los resultados obtenidos para la palabra *alguien*, que comparte las mismas características que la palabra *somebody* en el microlenguaje inglés.

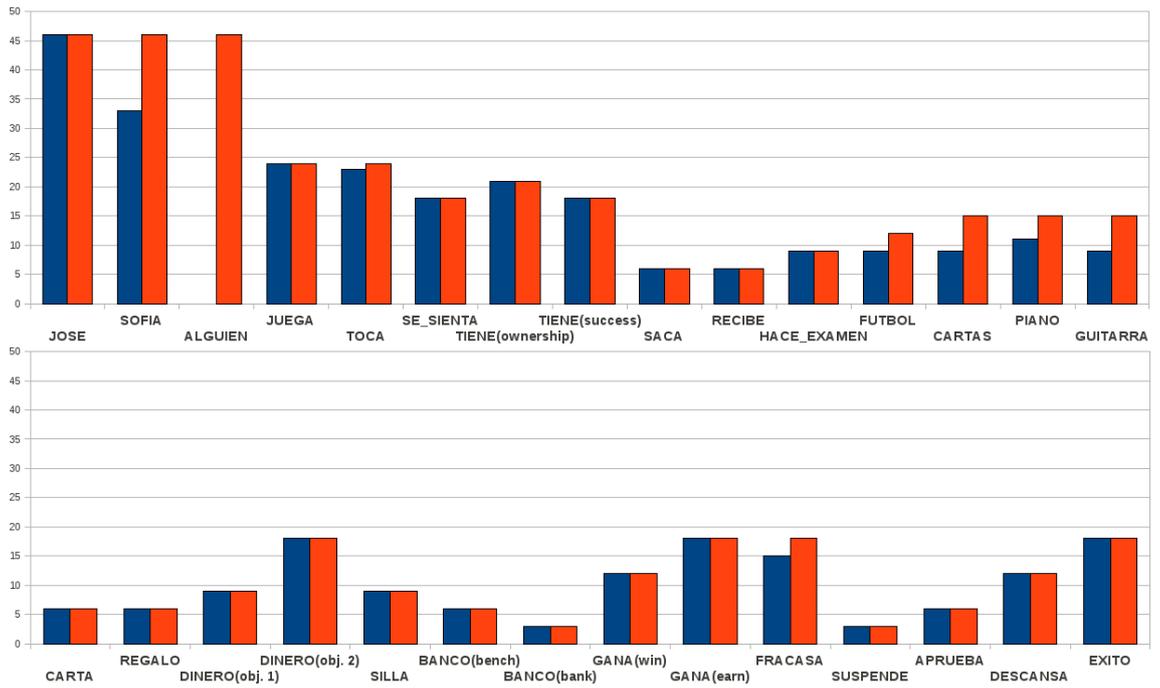


Figura 3.4: Resultados de acierto para cada palabra al traducir frases en español. El color rojo indica el total de apariciones en las frases de entrada, mientras el azul el número total de aciertos.

Una nueva medida de los resultados de traducción obtenidos anteriormente se presenta en las figuras 3.5 y 3.6. Esta nueva muestra cuenta con una distinción del error en dos casos: errores ligeros y errores graves. Se considera errores ligeros a aquellas traducciones que, a pesar de no haber devuelto la palabra correcta, han devuelto una traducción que corresponde al mismo subárbol conceptual. Por ejemplo, en el caso de *guitar*, se observa una fracción correspondiente a error ligero que resulta del hecho de que esta palabra fue traducida en algunos casos como *piano*, que es la otra palabra que forma parte del subárbol de conceptos “instrument”.

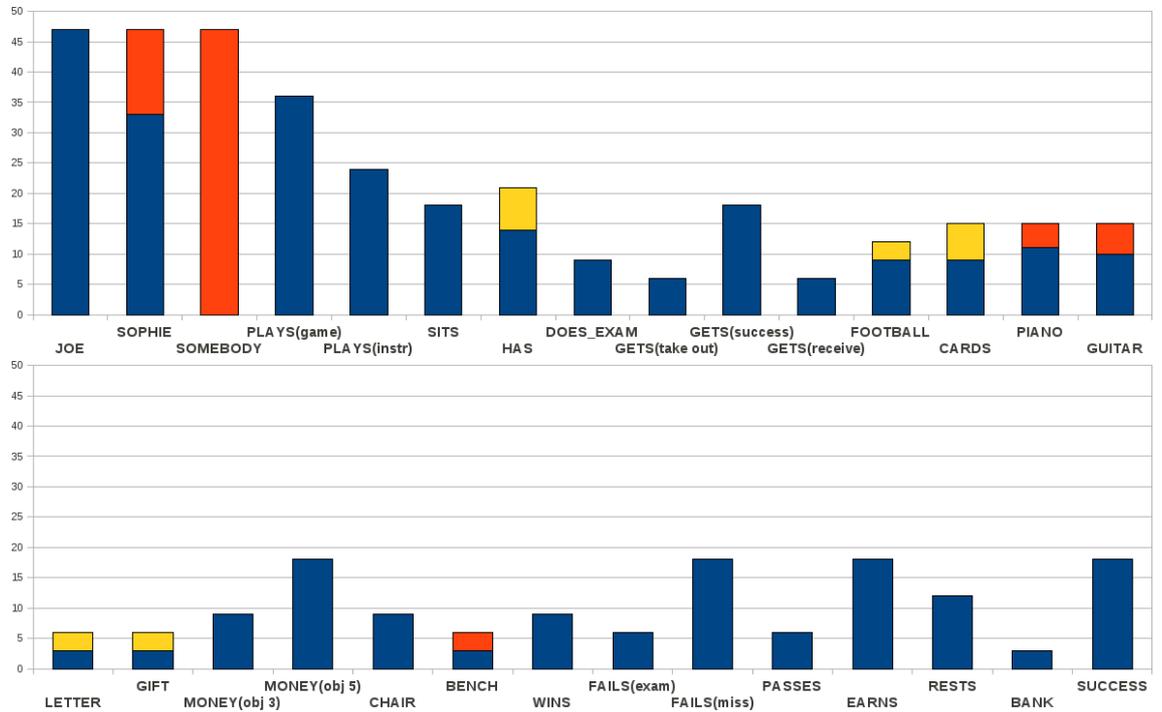


Figura 3.5: Resultados de acierto al traducir frases en inglés. El color azul indica el número total de aciertos, el color rojo el número de errores ligeros y el color amarillo son errores graves.

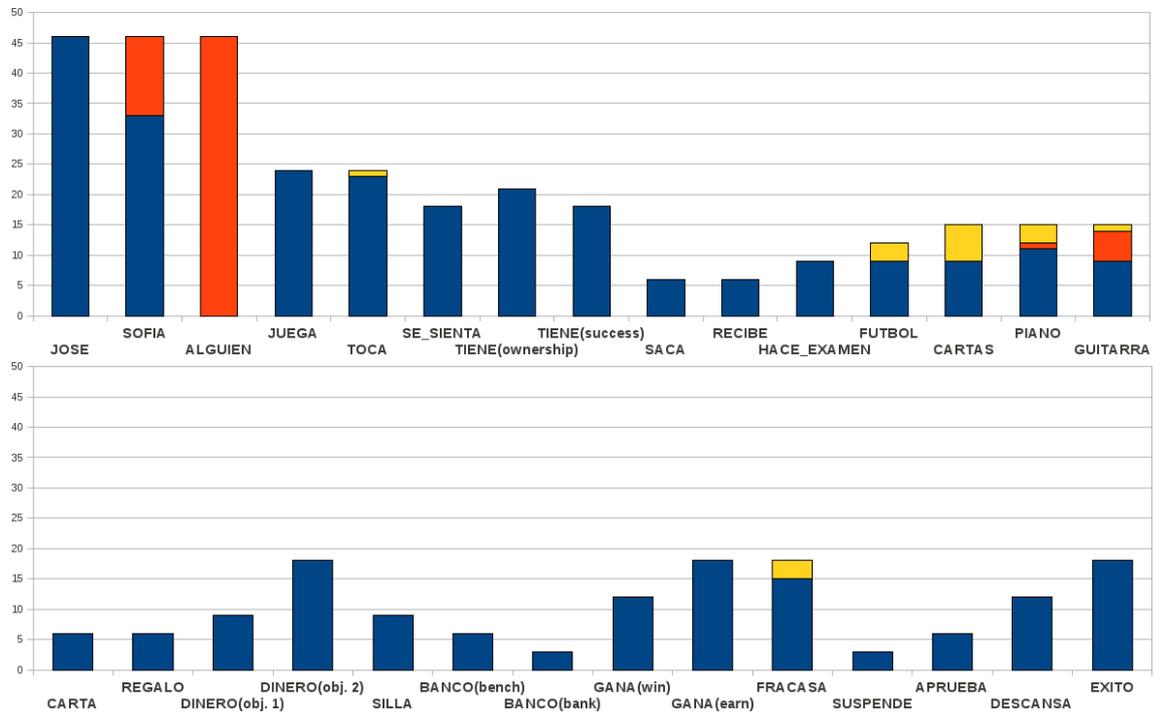


Figura 3.6: Resultados de acierto al traducir frases en español. El color azul indica el número total de aciertos, el color rojo el número de errores ligeros y el color amarillo son errores graves.

Capítulo 4

Conclusiones y discusión

Se ha descrito un modelo que posibilita la representación de situaciones en un micromundo mediante el uso de un mapa auto-organizativo y el procesamiento de frases que describen dichas situaciones utilizando una red recurrente simple. Posteriormente, se ha estudiado cómo un nuevo sistema basado en el primer modelo puede resultar adecuado para llevar a cabo tareas de “Machine Translation” entre dos idiomas distintos.

A continuación se destacan los componentes y características más interesantes del trabajo.

Formación de representaciones independiente del texto. El sistema desarrollado toma una frase como entrada del sistema y la transforma en la representación de la situación descrita, sin que exista ningún tipo de estructura proposicional intermedia. De esta manera, la representación formada como salida del modelo resulta ser independiente no sólo de algún tipo de estructura proposicional, sino también del lenguaje utilizado como entrada. En este trabajo, la representación situacional no se basa en la estructura lingüística presente en la entrada, sino que está basada en la experiencia del modelo sobre el mundo y su conocimiento sobre él.

Razonamiento lógico en el espacio de salida SOM. Se ha considerado cómo la representación utilizada puede ser útil para realizar razonamientos lógicos básicos al nivel de salida del mapa SOM como si de un diagrama Venn se tratara.

Comprensión de frases. Los resultados obtenidos en los experimentos de la primera de las partes del proyecto muestran que el modelo es capaz de llevar a cabo los dos tipos de generalización que se le exigen a un modelo de comprensión de frases según Frank [3]: es capaz de procesar con éxito nuevas descripciones de las situaciones en las que fue entrenado y, además,

es capaz de comprender descripciones de situaciones en las que no fue entrenado. De esta forma, el modelo no aprende simplemente la asociación entre palabras (o secuencia de palabras) y las situaciones, sino que aprende cómo *combinaciones* de palabras se refieren a *combinaciones* de situaciones.

Mapas de vocabulario. Además de la salida de representación situacional, dos nuevos mapas SOM son entrenados dando lugar a dos mapas de vocabulario diferentes del mismo tamaño. La organización de las palabras resultante en ambos mapas depende únicamente de los conceptos presentes en el micro-mundo y de los roles existentes en la definición de la gramática utilizada. Los mapas de palabras muestran cómo las palabras se han organizado en base a dos niveles diferentes, un primer nivel de organización por roles y un segundo nivel de organización conceptual, donde las palabras utilizadas para expresar conceptos y situaciones similares aparecen dispuestas más cerca en el mapa. Lo que resulta más interesante en esta organización conceptual de las palabras es que el modelo no es provisto desde el exterior de ningún índice subjetivo de relación entre cada una de las palabras de un lenguaje y los conceptos del mundo, ni siquiera de información sobre el enlace *palabra-concepto* que se pretende obtener para cada caso, sino que esta información es extraída de manera implícita por el sistema atendiendo al grado de influencia que una determinada palabra en una frase tiene sobre la aparición de una situación o concepto.

Proceso de traducción. El hecho de que exista una determinada representación situacional como salida en cada paso, sin necesidad de esperar a completar una frase, permite tener la posibilidad de obtener una traducción en todo momento desde que se presenta la primera palabra. La respuesta que el sistema ofrece a uno de estos casos en los que la palabra (o combinación de palabras) carece de un determinado contexto que permita desambiguar la traducción es devolver aquella traducción que tiene una mayor probabilidad de ocurrencia según la experiencia del sistema. Por otro lado, el hecho de que la representación situacional sea no dependiente del texto hace que el sistema sea potencialmente capaz de obtener, del mismo modo y gracias a la información de contexto, aquellos elementos conceptuales esenciales que subyacen en frases hechas, refranes o expresiones; aunque es cierto que esta capacidad habría de ser probada y los resultados habrían de ser analizados para comprobar si esta afirmación es cierta.

Abstracción y concreción de situaciones. Aquellas frases mencionando situaciones que tenían una alta frecuencia de aparición en el conjunto de entrenamiento y que fueron expresadas en el mismo idioma tienen una mayor probabilidad de provocar una salida más precisa como representación situa-

cional y, por lo tanto, una mejor traducción.

Sin embargo, este no ha sido un escenario habitual durante las pruebas realizadas en los experimentos y los porcentajes de acierto que se han obtenido son significativamente altos. La razón es que el modelo presentado es capaz de alcanzar cierto nivel de abstracción y concreción de situaciones al nivel del mapa SOM de salida: por un lado, el sistema excluye aquellos elementos o roles que puedan aparecer en el espacio de representación y que no son necesarios en la traducción y, por otro lado, al realizar una extracción de todos los roles requeridos en la traducción a partir del mapa de representación situacional, el sistema siempre ofrece una respuesta, incluso en el caso de que determinada palabra de entrada no esté presente en la experiencia del sistema (o su frecuencia de aparición sea baja), pudiendo dar lugar a los mencionados errores ligeros (ver 3.5, fig:espResult2).

4.1. Discusión

Antes de poder considerar que las redes neuronales simples son capaces de llevar a cabo aprendizaje de frases en general, el modelo debe ser aplicado a un lenguaje más complejo y dentro de un marco de micromundo que también lo sea. Además, es necesario tener en cuenta definiciones de microlenguaje y micromundo que sean más reales.

Actualmente, el modelo sólo puede actuar sobre una cantidad muy limitada de datos, donde el número de situaciones, frases y palabras es pequeño. Se requiere investigación futura para comprobar cómo la metodología utilizada escala para una aplicación real.

Los resultados de acierto obtenidos en los experimentos de traducción de frases entre los dos microlenguajes son altos, siendo todavía mayores en aquellos casos en los que la palabra a traducir tenía más de una correspondencia dentro del otro lenguaje (palabras ambiguas). Sin embargo, es cierto que una mayor cantidad de palabras en cada microlenguaje y una composición y gramática más complejas deberían ser tenidas en cuenta en posteriores experimentos. De cualquier manera, atendiendo a los resultados obtenidos, resulta prometedor realizar posteriores experimentos y comprobar qué resultados podrían ser obtenidos si se aplicara una organización conceptual y un método de asignación de relación concepto-palabra más complejos al modelo.

4.2. Trabajo Futuro

El trabajo desarrollado abre varias líneas de posible desarrollo futuro. Una de esas vías de estudio está relacionada con el conjunto de datos a utilizar por el modelo. Un modelo más completo de procesamiento de frases y traducción debería ser capaz de manejar un conjunto más grande y variado de datos de entrada. Además este conjunto debería contener ejemplos reales extraídos del mundo que nos rodea. Tener en cuenta un conjunto de datos de este tipo permitiría evaluar y estudiar el rendimiento del sistema en esos casos e investigar cuáles son las vías a seguir para conseguir un modelo escalable a la vez que funcional.

Por otro lado, este trabajo comparte con la gran mayoría de los trabajos actuales sobre procesamiento del lenguaje natural la problemática de la producción de frases. Esta es una de las tareas más complejas en NLP y, aunque no formaba parte de los objetivos a cubrir por este trabajo, un sistema de MT debería contar con las herramientas y componentes necesarios que permitan la producción de frases siguiendo las reglas gramaticales y proposicionales propias de cada lenguaje.

Bibliografía

- [1] DUCH, W. Neurocognitive informatics manifesto. In *Proceedings of IMS'09, the 8th International Conference on Information and Management Sciences*, pp. 264–282.
- [2] ELMAN, J. L. Finding structure in time. *Cognitive Science* 14, 2 (1990), 179–211.
- [3] FRANK, S. Sentence comprehension as the construction of a situational representation: A connectionist model. In *Proceedings of AMKLC'05, International Symposium on Adaptive Models of Knowledge, Language and Cognition* (Espoo, Finland, 2005), Helsinki University of Technology, pp. 27–33.
- [4] FRANK, S. Sentence comprehension without propositional structure. In *Modeling language, cognition and action* (New Jersey, London, 2005), World Scientific, pp. 119–128.
- [5] FRANK, S. L., KOPPEN, M., NOORDMAN, L. G. M., AND VONK, W. Modeling knowledge-based inferences in story comprehension. *Cognitive Science* 27, 6 (2003), 875–910.
- [6] FRENCH, R. M. A simple recurrent network model of bilingual memory. In *Proceedings of the Twentieth Annual Cognitive Science Society Conference* (Amsterdam, 1998), NJ:LEA, pp. 368–373.
- [7] GASSER, M. Connectionism and universals of second language acquisition michael, 1990.
- [8] HAM, F., AND KOSTANIC, I. *Principles of Neurocomputing for Science and Engineering*. McGraw–Hill, Singapore, 2001.
- [9] HARRIS, J. *Fuzzy Logic Applications in Engineering Science*. Springer, Dordrecht, 2006.

- [10] HONKELA, T., VIRPIOJA, S., AND VÄYRYNEN, J. Adaptive translation: Finding interlingual mappings using self-organizing maps. In *ICANN '08: Proceedings of the 18th international conference on Artificial Neural Networks, Part I* (Berlin, Heidelberg, 2008), Springer-Verlag, pp. 603–612.
- [11] III, M. R. M., AND MIIKKULAINEN, R. Incremental nonmonotonic sentence interpretation through semantic self-organization. Tech. Rep. AI08-12, Department of Computer Sciences, University of Texas at Austin.
- [12] KANGAS, J., TORKKOLA, K., AND KOKKONEN, M. Using SOMs as feature extractors for speech recognition. In *Proc. ICASSP-92, International Conference on Acoustics, Speech and Signal Processing* (Piscataway, NJ, 1992), IEEE Service Center.
- [13] KASKI, S., HONKELA, T., LAGUS, K., AND KOHONEN, T. Websom—self-organizing maps of document collections. *Neurocomputing* 21, 1 (1998), 101–117.
- [14] KOHONEN, T. *Self-Organizing Maps*. Springer, 2001.
- [15] KOHONEN, T., KASKI, S., LAGUS, K., SALOJÄRVI, J., NAD V. PAATERO, J. H., AND SAARELA, A. Self-organization of a massive document collection. *IEEE Transactions on Neural Networks* 11 (2000), 574–85.
- [16] LETOSA, J. R., AND HONKELA, T. Elementary logical reasoning in the som output space. In *ICANN (2)* (2010), K. I. Diamantaras, W. Duch, and L. S. Iliadis, Eds., vol. 6353 of *Lecture Notes in Computer Science*, Springer, pp. 432–437.
- [17] MAYBERRY, M. R., AND MIIKKULAINEN, R. SARDSRN: a neural network shift-reduce parser. In *IJCAI'99: Proceedings of the 16th international joint conference on Artificial intelligence* (San Francisco, CA, USA, 1999), Morgan Kaufmann, pp. 820–825.
- [18] MAYBERRY, III, M. R., AND MIIKKULAINEN, R. Incremental non-monotonic parsing through semantic self-organization. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (Boston, MA, 2003), Cognitive Science Society, pp. 798–803.
- [19] MAYBERRY III, M. R., AND MIIKKULAINEN, R. Using a sequential SOM to parse long-term dependencies. In *Proceedings of the 21st Annual*

- Meeting of the Cognitive Science Society (COGSCI-98)* (Hillsdale, NJ, 1999), Erlbaum.
- [20] MCCLELLAND, L., JOHN, M. F. S., AND MCCLELL, J. L. Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence* 46 (1990), 217–257.
- [21] MIIKKULAINEN, R. *DISCERN: A Distributed Artificial Neural Network Model of Script Processing and Memory*. PhD thesis, Computer Science Department, University of California, Los Angeles, 1990. (Tech. Rep UCLA-AI-90-05).
- [22] MIIKKULAINEN, R. *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*. MIT Press, Cambridge, MA, 1993.
- [23] NOELLE, D. C., AND COTTRELL, G. W. A connectionist model of instruction following. In *Moore, J. D., and Lehman, J. F., editors, Proceedings of the 17th Annual Conference of the Cognitive Science Society* (Pittsburg, 1995), Erlbaum, pp. 369–374.
- [24] PRESS, H. U., Ed. *Mental Models* (Cambridge, 1983).
- [25] ROHDE, D. L. T. *A connectionist model of sentence comprehension and production*. PhD thesis, Pittsburgh, PA, USA, 2002. Chair-Plaut, David C.
- [26] STEELS, L. Social language learning. In *Tokoro, M. and L. Steels, editors, The Future of Learning* (Amsterdam, 2003), IOS Press, pp. 133–162.
- [27] ZADEH, L. A. Fuzzy sets. *Information and Control* 8 (1965), 338–353.

Apéndice A

Métodos

Algoritmo Self-Organizing Map

El mapa auto-organizativo (Self-Organizing Map) es un tipo de red neuronal artificial en el que el aprendizaje es llamado *competitive, unsupervised* o *self-organizing*. Por medio del algoritmo SOM, la red desarrolla un clustering de pequeñas dimensiones (típicamente dos dimensiones) del espacio de entrada de las muestras de entrenamiento a la vez que preserva las propiedades topológicas de ese espacio por medio del mapeo de patrones de entrada similares a unidades cercanas en el mapa de salida.

Estas unidades del Self-Organizing Map, llamadas nodos o neuronas, están dispuestas normalmente en un espacio regular en una cuadrícula rectangular o hexagonal, y cada una de ellas está asociada con un vector de pesos de la misma dimension que los vectores de datos de entrada.

Durante el algoritmo auto-organizativo, el nodo del mapa cuyos pesos se encuentran más cercanos a un patrón de entrada es activado. Entonces, estos pesos y los pesos de los nodos de alrededor son adaptados hacia el patrón de activación de la entrada, dirigiéndose a una situación en la que los nodos cercanos se hacen similares y responden a patrones similares de entrada.

Algoritmo de Entrenamiento

El primer paso en el algoritmo de entrenamiento pasa por inicializar los vectores de pesos ya sea de manera aleatoria o seleccionando un conjunto de pesos que refleje algún tipo de conocimiento *a priori* sobre los datos de entrada.

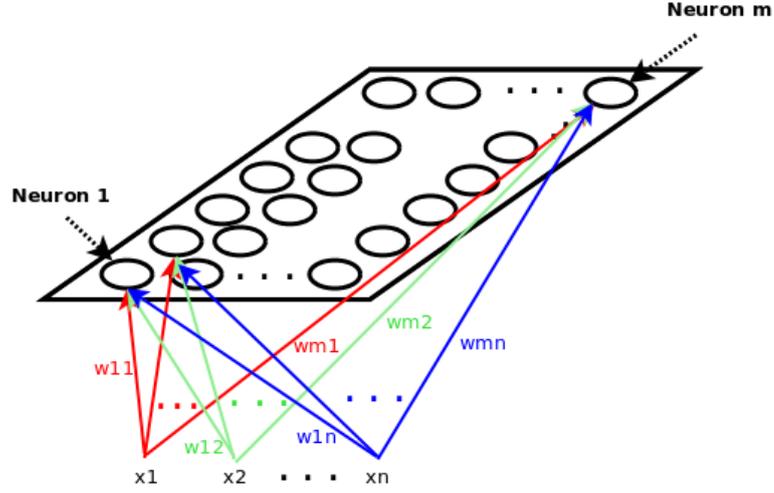


Figura A.1: Arquitectura SOM.

Las entradas de la red mostrada en la figura A.1 pueden ser escritas en forma de vector como

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T \quad (\text{A.1})$$

y el vector de pesos sináptico de la neurona i es dado por

$$\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T \quad i = 1, 2, \dots, m \quad (\text{A.2})$$

donde m es el número total de neuronas en el vector de salida. La mejor coincidencia del vector de entrada x con el vector de pesos sináptico \mathbf{w}_i está determinado por

$$q(\mathbf{x}) = \min_{\forall_i} \|\mathbf{x} - \mathbf{w}_i\|_2 \quad i = 1, 2, \dots, m \quad (\text{A.3})$$

donde $q(\mathbf{x})$ es el índice del vector de salida que identifica específicamente la *neurona ganadora*, y $\|\bullet\|_2$ es la norma euclídea. La respuesta podría ser también el vector de pesos sináptico más cercano a la entrada en vez de un índice de posición. El siguiente paso en el algoritmo consiste en actualizar el vector de pesos sináptico asociado con la neurona ganadora y las neuronas que se encuentran dentro de un espacio de proximidad definido. La regla de aprendizaje viene dada por

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \eta_{qi}(k)[\mathbf{x}(k) - \mathbf{w}_i(k)] \quad (\text{A.4})$$

donde

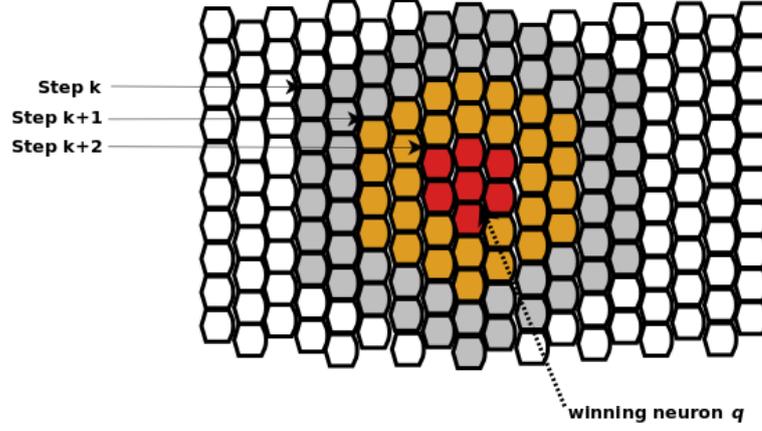


Figura A.2: Celdas vecinas, mostrando un decrecimiento monótono. En el paso k , la función $N_q(k)$ cubre las celdas vecinas coloreadas en gris, naranja y rojo. Para el siguiente paso, $k + 1$, el conjunto de vecinos cubre las celdas naranjas y rojas y, finalmente, para el paso $k + 2$ el conjunto ha sido reducido a las celdas rojas.

$$\eta_{qi}(k) = \begin{cases} \mu(k) & \text{en el conjunto vecindario } N_q \text{ para la neurona ganadora } q \\ & \text{donde } 0 < \mu(k) < 1, \text{ el parámetro de ratio de aprendizaje} \\ & \text{que debería decrecer con el tiempo} \\ 0 & \text{fuera de } N_q \end{cases} \quad (\text{A.5})$$

es una función de proximidad. Teniendo en cuenta las ecuaciones A.4 y A.5, la regla de aprendizaje puede ser escrita como

$$\mathbf{w}_i(k+1) = \begin{cases} \mathbf{w}_i(k) + \mu(k)[\mathbf{x}(k) - \mathbf{w}_i(k)] & \text{if } i \in N_q(k) \\ \mathbf{w}_i(k) & \text{if } i \notin N_q(k) \end{cases} \quad (\text{A.6})$$

donde el *parámetro de ratio de aprendizaje*, $\mu(k) \in (0, 1)$, y $N_q(k)$ es una función del índice de tiempo discreto k . Se ha demostrado que es beneficioso hacer que el conjunto de proximidad $N_q(k)$ sea relativamente ancho al comienzo del entrenamiento para después reducirlo monótonamente con el tiempo. La figura A.2 muestra un decremento monótono del conjunto de proximidad.

Durante el proceso de aprendizaje hay dos fases relacionadas: la *fase de ordenación* y la *fase de convergencia*. Durante la *fase de ordenación* (proceso

de aprendizaje inicial) la ordenación topológica de los vectores de pesos es llevada a cabo. La segunda fase, la *fase de convergencia*, es generalmente la parte más larga del aprendizaje, y es durante esta fase cuando el algoritmo tiene que llevar a cabo los ajustes más precisos del mapa.

Trabajo Relacionado

El algoritmo Self-Organizing Map ha sido usado antes en muchas tareas relacionadas con el Procesamiento del Lenguaje Natural (NLP) y la traducción máquina (MT). Algunos ejemplos de estos trabajos incluyen minería de datos [15], [13], reconocimiento del habla [12], o *parsing* [19].

En [21], un doble mapa es usado como un diccionario que enlaza por medio de aprendizaje no supervisado el símbolo ortográfico de una palabra con su representación semántica. Cada uno de los dos mapas del diccionario están organizados independientemente. De esta manera, en el mapa físico, las representaciones de las palabras que son más parecidas desde un punto de vista léxico aparecen mapeadas unas cerca de las otras; mientras el mapa semántico refleja una división de tres áreas principales: verbos, objetos animados y objetos inanimados. La asociación resultante entre los dos mapas es un mapeo muchos a muchos entre la representación física y la semántica de las palabras, permitiendo una conexión bidireccional entre palabras físicas que pueden estar relacionadas con varias representaciones semánticas y viceversa.

Posteriormente, Mayberry y Miikkulainen [18, 11] usan una versión del SOM para su modelo INSOMNet para desarrollar representaciones semánticas a partir de un corpus. INSOMNet es un modelo supervisado de comprensión y *parsing*. Este sistema interpreta frases de manera no monótona, generando expectativas y revisándolas, aplicando futuras entradas basadas en semántica, y coactivando propiamente múltiples interpretaciones de expresiones [11]. Su arquitectura está formada por tres módulos operacionales principales: *sequence processor*, *semantic frame encoder and decoder* y *frame selector*. El primero de ellos incluye el mapa SARDNet ([19]), que es un mapa auto-organizativo para resolver el problema de memoria a largo plazo del SRN. Un segundo mapa auto-organizativo es usado, el *Frame Node Indicator Map*, que es parte del módulo *frame selector*. Sus resultados muestran que INSOMNet aprende a representar dependencias semánticas de manera precisa y generaliza a estructuras originales.

El Self-Organizing Map ha sido usado también en el estudio de Machine Translation como fue mencionado en el primer párrafo de esta sección. En [10], un mapa auto-organizativo sirve de espacio conceptual en el que surgen ciertas clases de representación interlingual entre dos lenguajes (alemán e inglés, en este caso) [10]. Se llevan a cabo dos experimentos. En el primero de ellos, una colección de palabras de los lenguajes inglés y alemán son considerados en diferentes contextos situacionales de la vida real. Este primer experimento muestra cómo los contextos de situación están organizados en el mapa de manera que el mapa está dividido en diferentes áreas conceptuales y, al mismo tiempo, cómo un mapeo entre las palabras de ambos lenguajes ha sido producido.

Este trabajo destaca la idea de que diferentes lenguajes conceptualizan la realidad de manera diferente [26].

Finalmente, el trabajo de Frank ha sido mencionado ya que este trabajo está próximamente relacionado a la aproximación de Frank. Comenzando con un modelo básicamente centrado en el uso de mapas auto-organizativos para llevar a cabo representación del conocimiento basado en la comprensión de situaciones de una historia creada por la combinación de proposiciones básicas y el estudio sobre cómo extraer inferencias de esta representación [5], Frank añade un modelo de comprensión de frases [4, 3] basado en un microlenguaje que pretende aprender cómo relacionar un conjunto de frases de entrada con las situaciones del micromundo representadas sobre el mapa SOM, y desempeñar cierto aprendizaje de abstracción en la comprensión de frases.

Simple Recurrent Neural Network

La Red Recurrente Simple (SRN; [2]) es una red neuronal *feedforward* de una sola capa oculta. Sin embargo, la salida de esta capa oculta es retropropagada a la entrada de la red por medio de una *capa de contexto* (o *capa oculta previa*), que es un ensamblaje extra de neuronas. La función de la capa oculta previa es replicar la salida de la capa oculta en el paso anterior, sirviendo, de esta manera, como contexto temporal para la entrada en el siguiente paso.

La SRN ofrece un modo de implementar un tipo de sistema que potencialmente almacena memoria. De hecho, en el marco de este trabajo, este tipo de red permite distinguir cuál es la situación correcta especificada por la frase incluso si el mismo vector de entrada (palabra de la frase, w_i) en el paso k puede aparecer en diferentes frases refiriéndose -cada frase - a una situación

diferente. Esto es posible teniendo en cuenta que las palabras anteriores de la frase han sido ya presentadas a la red y la capa oculta representa el *contexto* apropiado para la palabra w_i .

Sin embargo, la SRN tiene el problema de que la información representada por la capa de contexto se degrada a medida que cada nueva entrada es presentada a la red. Este problema de memoria se debe al hecho de que la longitud de la capa de contexto es fija, y la información de la señal recibida de pasos anteriores puede decrecer para adaptarse a la llegada de señales más recientes.

La figura A.3 muestra la arquitectura genérica para una red recurrente simple. El vector de entrada $\mathbf{x}(k)$ es presentado a la red y propagado a través de la *capa oculta*, que recibe en el mismo paso la señal de la *capa de contexto*, $\mathbf{x}^c(k)$, estableciendo de esta manera un contexto para el vector de entrada $\mathbf{x}(k)$. La salida de la *capa oculta* es, entonces, propagada a la capa de salida y, al mismo tiempo, una copia de esta salida de la *capa oculta* es retropropagada a través de las unidades de contexto para preparar el contexto del siguiente vector de entrada.

Trabajo Relacionado

Desde su aparición, las redes SRN han sido una de las herramientas más usadas para sistemas conexionistas de procesamiento de lenguaje natural [18]. En las publicaciones de St. John y McClelland [20] y McClelland, St. John, and Taraban, la mitad de la entrada del modelo es una red neuronal recurrente simple cuyo objetivo es aprender a compilar una representación de un solo mensaje a partir de una secuencia de componentes de una frase. Los componentes de la frase pueden ser una simple frase sustantiva, una frase preposicional, o una verbal [25]. El modelo exhibe características interesantes, teniendo la habilidad de hacer uso tanto de las pistas sintácticas como semánticas para inferir significado, además de ser capaz de inferir constituyentes no presentes. Por ejemplo, que la acción “comer sopa” se realiza probablemente con una cuchara.

Otros modelos inspirados en el trabajo de St. John y McClelland [20] son el trabajo de Noelle y Cottrell [23] y el modelo conexionista de producción y comprensión de frases (CSCP) de Rohde [25]. En el trabajo de Noelle y Cottrell la suposición de que la mayoría del lenguaje es aprendido durante la ejecución de diferentes tareas está presente en el sentido en que están interesados en la capacidad de realizar una tarea después de recibir algunas

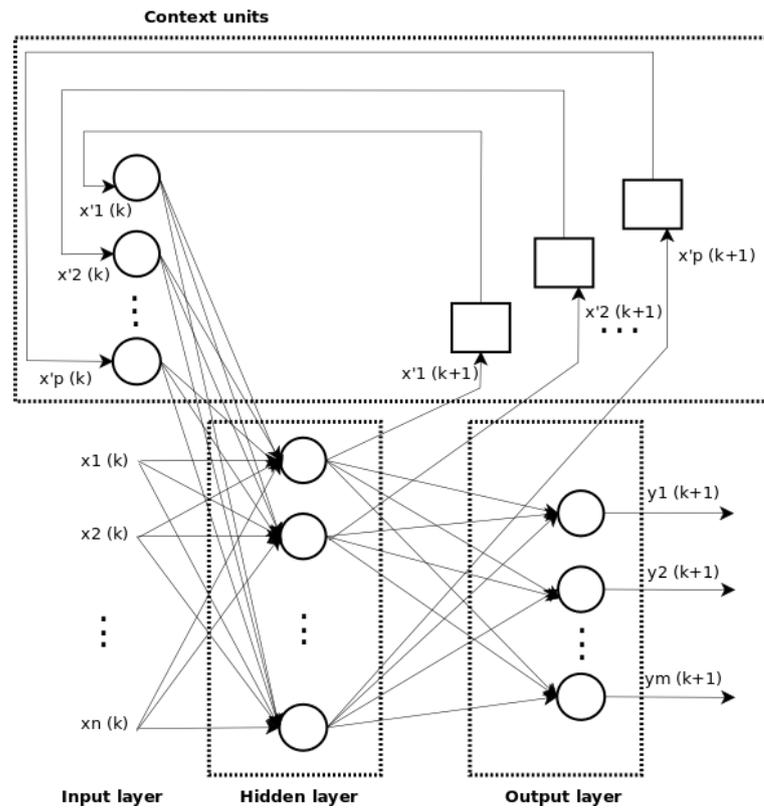


Figura A.3: Arquitectura SRN.

instrucciones de cómo hacerlo. El componente *plan* de la red recibe instrucciones en el tiempo y produce un plan que guía la acción de la red.

El CSCP es también un sistema basado en la SRN, creado, esta vez, para llevar a cabo tanto comprensión como producción de frases. En este modelo se usa una red recurrente simple para aprender a codificar la estructura proposicional de significado en las frases de entrada en mensajes estáticos comprimidos. Además, al mismo tiempo que el sistema realiza la tarea de comprensión, una tarea de predicción usando la SRN, que es la base del modelo para aprender a producir lenguaje, es llevada a cabo. Argumentando que la producción es, en un sentido, la habilidad para predecir lo que un hablante normal diría para transmitir el intencionado mensaje (si ignoramos los requerimientos físicos) [25], la tarea de producción consiste básicamente en decodificar el mensaje comprimido previamente mencionado en su proposición correspondiente. En el momento en el que la frase es presentada al modelo y el mensaje es representado, el sistema predice cuál es la siguiente palabra de la frase de entrada y la produce como salida.

La Red Recurrente Simple en este Trabajo

Como ha sido comentado, la red recurrente simple (SRN; [2]) ha sido utilizada en trabajos anteriores para realizar tareas relacionadas con NLP como principal herramienta o como componente en sistemas de análisis.

El esquema de SRN utilizado se muestra en la figura A.4. La capa de entrada de la red consiste en 13 unidades, una unidad por cada palabra en el *microlenguaje*. Esta disposición se corresponde con la representación localista (figura A.4). La figura muestra un vector de color blanco donde sólo una de las componentes, w_i , está activa en cada representación ($w_i = 1$) y el resto de los elementos es igual a 0. De esta manera, cada una de las palabras del *microlenguaje* está relacionada unequivocamente con un índice del vector. Siguiendo esta idea, una frase en el *microlenguaje* se presenta a la red como la secuencia ordenada de palabras que la forman.

El vector de entrada junto con las señales de la capa de contexto son propagadas a la capa oculta de 50 unidades, donde se forma una representación distribuida interna. Los valores de la capa oculta han sido estudiados en otros trabajos relacionados con NLP en la búsqueda de cualquier patrón creado por la red de manera independiente durante el entrenamiento que permita encontrar características gramaticales.

En [6], Robert M. French entrena una red recurrente simple para dos microlenguajes y las representaciones de las unidades de la capa oculta son

estudiadas. La entrada de la red se compone de frases en inglés y francés ordenadas aleatoriamente, de manera que no hay marcadores para delimitar el comienzo ni el fin del cambio de lenguaje. El objetivo de la SRN esta vez es predecir la siguiente palabra de la frase que está siendo presentada. Después del entrenamiento, se calcula el valor medio de las 32 unidades de la red para ambos lenguajes. Los experimentos muestran que estas representaciones altamente distribuidas aparecen agrupadas no sólo por estructuras gramaticales, sino también para cada uno de los dos lenguajes (llamados *Alpha* y *Beta* en el experimento).

En el modelo de comprensión de Stephan L. Frank [4], que está relacionado con el implementado en este trabajo, un conjunto de 32 tripleteas de frases son utilizadas para medir las distancias medias de cada tripletea en los siguientes términos: distancia textual, distancia proposicional y distancia situacional. Cada tripletea es elegida de manera que existen dos frases, *A* y *B*, cercanas a nivel textual, entonces las frases *A* y *C* están relacionadas proposicionalmente y, finalmente, *B* y *C* están relacionadas situacionalmente, es decir, estas frases se refieren situaciones similares incluso aunque la representación textual de ambas frases no sea similar. Los resultados muestran que la distancia más larga se produce al nivel proposicional; resultando en un modelo que difiere de lo que es tradicionalmente pensado sobre la necesidad de estructuras proposicionales intermedias para constituir la base textual [4].

Para continuar con la descripción del esquema en este trabajo, la última capa (o capa de salida) todavía tiene que ser mencionada. Esta capa ha de reflejar en cada paso la representación distribuida de la situación en el micromundo referida por la frase de entrada. La representación de la situación mostrada en la salida puede variar a medida que las palabras de la frase son presentadas a la red para completarla. De hecho, en la frase 'LC performs a concert in Europe', la situación representada después de las primeras cuatro palabras ('LC performs a concert') es diferente que la representada después de que la frase completa es presentada. La capa de salida está formada por 150 unidades, $s_i \in [0, 1]$, cada una de ellas correspondiendo a una celda del espacio de salida del SOM.

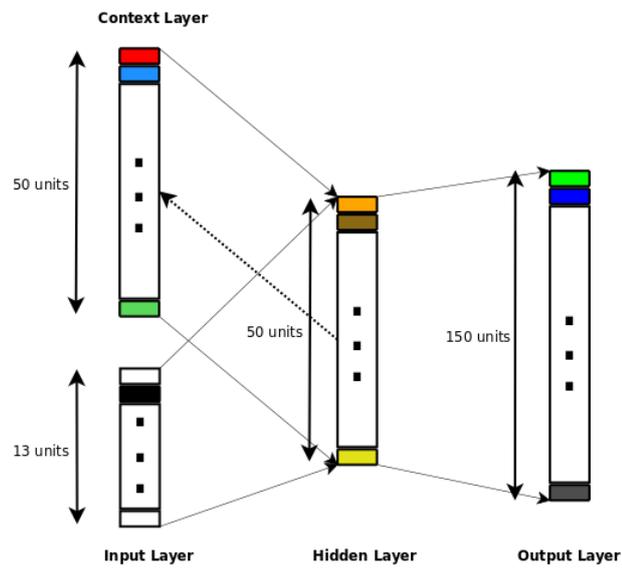


Figura A.4: Arquitectura SRN en este trabajo. La capa de entrada muestra una representación localista de cada palabra, mientras los vectores coloreados expresan representación distribuida.

Ap ndice B

Elementary Logical Reasoning in the SOM Output Space

Elementary Logical Reasoning in the SOM Output Space^{*}

Jorge Ramón Letosa^{**} and Timo Honkela

Adaptive Informatics Research Centre
Aalto University of Science and Technology
P.O. Box 15400, FI-00076 Aalto

Abstract. In this paper, we consider how to represent world knowledge using the self-organizing map (SOM), how to use a simple recurrent network (SRN) to device sentence comprehension, and how to use the SOM output space to represent situations and facilitate grounded logical reasoning.

1 Introduction

The self-organizing map (SOM) is a well known neural network algorithm. The map is typically a 2-dimensional array, the nodes of which become specifically tuned to various input signal patterns or classes of patterns in an orderly fashion [1]. The SOM provides a means for modeling concept formation and symbol grounding [2, 3]. It also can be used for implicit representation of conceptual hierarchies [1]. In this paper, we focus on the question how to conduct basic logical reasoning within the SOM framework. Moreover, we wish to connect the level of representing knowledge propositionally with its representation in the form of natural language sentences. The framework would also allow multimodal grounding of the knowledge (see e.g. [4]) but we do not address the issue here.

In a *related work*, Duch has provided an overall view and motivation on neurocognitive modeling of linguistic processing [5]. In relation to conceptual modeling, he states that each node in a neurally plausible semantic network is a neural circuit, with similarities and associations between concepts resulting from sharing some common elements or mutual activations that are responsible for semantic priming [5]. Consistent with this line of thought, Miikkulainen developed already some time ago a system for story comprehension using the SOM as a conceptual memory [6]. Later, the model has been extended to include recurrent processing of sentence structures [7]. In a recent work, Mayberry and Miikkulainen have introduced a new connectionist model called InSomNet [8]. Their results show that InSomNet learns to represent semantic dependencies accurately and generalizes to novel structures. The InSomNet system interprets sentences nonmonotonically, generating expectations and revising them, priming

^{*} The work has been supported by the Academy of Finland.

^{**} Visiting from the University of Zaragoza, Spain

future inputs based on semantics, and properly coactivating multiple interpretations of expressions [8].

Frank has presented a model in which basic logical operations can be represented as a certain kind of Venn diagram manipulations on the SOM surface [9]. Our current work is closely related to Frank's approach. We verify his experimental results, provide some extensions, and build further the underlying theoretical framework. Starting with a model basically focused in the use of self-organizing maps to carry out knowledge representation based on the comprehension of situations from a story created from the combination of basic propositions and the study about how to extract inferences from this representation [10], Frank adds a sentence comprehension model [11, 9] based on a microlanguage that intends to learn how to relate a set of input sentences with the microworld situations represented onto a SOM map, and accomplish some kind of abstraction learning in the sentence comprehension.

As the *main objectives* in this paper, we consider how to represent propositionally defined situations using the SOM, how to represent the processing and structural representation of related sentences using recurrent neural networks, and how to evaluate the comprehensibility of the sentences in the framework of this methodology. We also consider how logical reasoning can be conducted using the SOM output space as grounding for the propositional elements. We present the underlying theory and methodology as well as an illuminative experiment.

2 Methods

The objective of this work is threefold: (1) to represent world knowledge using the self-organizing map (SOM), (2) to use a simple recurrent network (SRN) to device sentence comprehension, and (3) to use the output space to represent situations and facilitate grounded logical reasoning. We will apply set theory and fuzzy set theory [12], clause logic, neural networks (specifically simple recurrent neural networks [13] and the self-organizing map [1]).

The degree of fit between an input value and a particular prototype in the SOM is interpreted as an membership value in the fuzzy set theory. We also assume that the proportional size of the distribution on the map related to a specific input approximates the probability of this input. Furthermore, we assume that the distributions on the map can be interpreted as fuzzy Venn diagrams allowing for basic set-theoretic operations (see e.g. [14]). Venn diagrams show possible logical relations among a finite collection of sets.

Simple recurrent network is a variant multilayer perceptrons. In a simple recurrent network, a set of context units is added to the three-layer network. The connections from the hidden layer to context units allow the network to model time-dependent phenomena and perform tasks like sequence prediction. [13]

2.1 World Knowledge Representation

In order to implement world knowledge, we shall begin by constructing a defined microworld framework in which a set of situations takes place. After defining the

microworld and its constraints, we explain how the microworld knowledge can be represented by means of the help of the Self-Organizing Maps.

Following the model presented in [10], the knowledge of a microworld is learned implicitly by means of training a set of example situations that fulfill the microworld definition; each one of them is created from combining the events such as presented in Table 1. Thus, the SOM output of the model is expected to reflect the inherent regularities, constraints and the *a priori* probabilities of occurrences of the concrete events within the microworld. This serves as a kind of “experience” the system has after being trained with a set of microworld example situations.

Table 1. Examples of the basic events in the microworld.

No.	Name	Meaning
1	70s	In the 70s decade.
2	80s	In the 80s decade.
3	lcSuccess	LC gets success.
..
15	bdHome	BD is at the home country.
16	bdEurope	BD is in Europe.
17	accompanied	He is/they are accompanied.

After training, a situation in the microworld is represented by a high-dimensional vector of $n = 150$ components (a SOM map). This implementation in which a situation is contained in a vector of fixed size of n elements allows a representation in which n is independent of the size of the example set, although, at the same time, by reducing the dimensionality from k , size of the example set, 227 in this case, to $n = 150$, some of the information contained in the previous example situations set may be lost. A situation vector is in the form $s(p) = (s_1(p), s_2(p), \dots, s_n(p))$, where every $s_i(p)$ has a value between 0 and 1 for every event p (p can, in fact, be any combination of events) that indicates the extent to which the component, or more precisely, the cell in the SOM, is a part of the representation of p ([9]). As it can be observed in Fig. 1, the pattern representing “*lcHome*” overlaps with the pattern in which “*lcWrites*”, showing that if LC is writing a book, LC is at the home country.

In general, a situation can be represented on a map so that it is possible to obtain the probability of occurrence of such situation and any combination of situations by means of fuzzy set theory. Taking the example shown in Fig. 2 where p and q are situations with a probability of 0.5 each one in the microworld and taking into account the equations $s_i(\neg p) = 1 - s_i(p)$ and $s_i(p \wedge q) = s_i(p)s_i(q)$ every logical combination of situations can be represented.

Since there is no one-to-one correspondence between propositions and dimensions in the taken representation of situations, we use *belief values* in order to figure out the results that are given as situation vectors. Let $X = (x_1, x_2, \dots, x_n)$

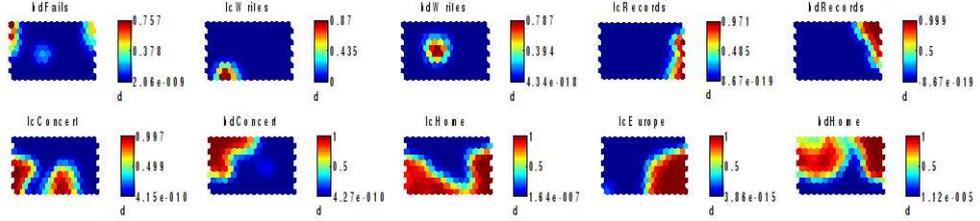


Fig. 1. A subset of the component planes of a situation map.

be a situation vector, with n as the number of situation-space dimensions. As a result of training the SOM, the probability (the probability estimate is approximated) that situation X occurs in the microworld equals the fraction of the map it covers ([9]). The belief values τ for a situation X and for a situation p in X are defined as:

$$\tau(X) = \frac{1}{n} \sum_i x_i. \quad \tau(p|X) = \frac{\sum_i s_i(p) x_i}{\sum_i x_i}. \quad (1)$$

2.2 Sentence Comprehension

Next, we convert sentences from a microlanguage to the representation of the corresponding microworld situation.

The *microlanguage* used in the experiments consists of 13 words: *LC*, *BD*, *and*, *is*, *gets_success*, *fails*, *performs_concert*, *writes_book*, *records_CD*, *at_home*, *in_Europe*, *single*, *accompanied*. With these words and following the rules of the grammar in Table 2, 396 different sentences can be obtained by associating each of them with a microworld situation.

Microlanguage sentences are transformed into situational vector representations by training a simple recurrent neural network [13]. The input layer of the network consists of 13 units, one for each word in the microlanguage. The hidden layer has 50 units and the output layer 150 units, one for each dimension of the situational space.

The words of a sentence are processed one-by-one, so only one of the input units is active at any moment. This way, the output of the network can be

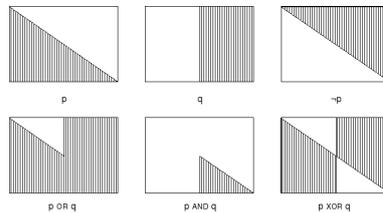


Fig. 2. Basic logical reasoning at the SOM output level.

Table 2. Grammar of the microlanguage.

S	→ NP VP
NP	→ LC BD LC and BD BD and LC
VP	→ Action [Place and is State and Result] → is Place [and Action and State and Result] → is State [and Action Place and Result] → Result [and Action Place and is State]
Action	→ writes_book performs_concert records_CD
Place	→ at_home in_Europe at_home and in_Europe in_Europe and at_home
State	→ single accompanied
Result	→ gets_success fails

observed at any moment and it is possible to analyze how the representation of the situations gets defined as the sentence gets completed.

Belief values are useful in measuring the performance of the model after the network has been trained. The belief value of a situation $X(p)$, represented by the output for a certain sentence, should be larger than the a priori belief value of the situation p corresponding to that sentence:

$$compr(p) = \frac{\tau(p|X(p)) - \tau(p)}{\tau(p|p) - \tau(p)}. \quad (2)$$

When $\tau(p|X(p)) > \tau(p)$, the comprehension score is higher than 0 and the network reflects a comprehension of the sentence, having the ideal case when $\tau(p|X(p)) = \tau(p|p)$, and the score is 1. On the other hand, when the score is negative the network is misunderstanding the sentence and there is no comprehension when the result equals 0.

3 Experiments

In the following, we describe the experiments conducted in this work including both the formation of the self-organizing map of the situations and simple recurrent network model of the sentences describing the situations in a microworld.

The example situations set constitutes the input of the SOM and consists of 227 example situations that follow the microworld constraints so that every situation in this set is obtained by one or more propositions combined. Each of the input example situations is a vector $\mu = (\mu_1, \mu_2, \dots, \mu_m)$, $m = 17$, which is the number of events in the microworld, so that a component μ_i is equal to 1 if it is the case of the example situation or 0 if it is not.

The SRN network is trained with a set of 368 sentences that are presented to the network randomly. The rest of the sentences are not trained and they are used as a test set. After training and calculating the comprehension scores for the sentences present in the training set, the percentage of comprehended sentences is 89.7% and the mean of the score is 0.3931. In the other hand, a set

of 20 sentences not presented to the network during the training but referring to situations in the microworld that were already present in the training set are considered as well, obtaining a score mean of 0.3330 within the 18 of the 20 sentences that were comprehended. Finally, a set of 8 sentences not presented describing new situations resulted in a score mean of 0.3799 and 7 of this set of 8 sentences had a comprehension score above 0. Furthermore, the results show how the short sentences lead to higher comprehension scores than the long ones.

4 Conclusions

In this paper, we have described an approach which enables representation of situations in a microworld using the self-organizing map algorithm and processing of sentences that describe the situations using simple recurrent network. We have also considered how to conduct basic logical reasoning at the output level of the self-organizing map in a Venn diagram like manner. The approach seems to facilitate well a grounded approach for modeling story comprehension. The present study is based on a microworld in which the number of different situations and sentences is small. Future research is needed to test how well the methodology scales up to large real-world applications.

References

1. Kohonen, T.: Self-Organizing Maps. Springer (2001)
2. Gärdenfors, P.: Conceptual spaces: The Geometry of Thought. MIT Press (2000)
3. Honkela, T.: Self-Organizing Maps in Symbol Processing. In: Hybrid Neural Systems. Springer (2000) 348–362
4. Harnad, S.: The symbol grounding problem. *Phys. D* **42**(1-3) (1990) 335–346
5. Duch, W.: Neurocognitive informatics manifesto. In: Proc. of IMS'09. 264–282
6. Miikkulainen, R.: Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory. MIT Press, Cambridge, MA (1993)
7. Mayberry, M.R., Miikkulainen, R.: SARDSRN: a neural network shift-reduce parser. In: Proc. of IJCAI'99, San Francisco, CA, Morgan Kaufmann (1999) 820–825
8. Mayberry, M.R., Miikkulainen, R.: Incremental nonmonotonic sentence interpretation through semantic self-organization. Technical Report AI08-12, Department of Computer Sciences, University of Texas at Austin
9. Frank, S.L.: Sentence comprehension as the construction of a situational representation: A connectionist model. In: Proceedings of AMKLC'05, International Symposium on Adaptive Models of Knowledge, Language and Cognition, Espoo, Finland, Helsinki University of Technology (2005) 27–33
10. Frank, S.L., Koppen, M., Noordman, L.G.M., Vonk, W.: Modeling knowledge-based inferences in story comprehension. *Cognitive Science* **27**(6) (2003) 875–910
11. Frank, S.L.: Sentence comprehension without propositional structure. In: Modeling language, cognition and action, New Jersey, World Scientific (2005) 119–128
12. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8** (1965) 338–353
13. Elman, J.L.: Finding structure in time. *Cognitive Science* **14**(2) (1990) 179–211
14. Harris, J.: Fuzzy Logic Applications in Engineering Science. Springer, Dordrecht (2006)