Clasificación Fonética Independiente del Locutor usando una Jerarquía de RNAs Especializadas

Hernando Silva Varela y Valentín Cardeñoso Payo *

Departamento de Informática (ETIT)
Universidad de Valladolid, España

hernando@infor.uva.es

Resumen:

Se describe una técnica para la clasificación fonética independiente del locutor mediante el uso de Redes Neuronales Artificiales (RNAs) especializadas; la técnica propuesta se basa en el principio "dividir para vencer" al utilizar una jerarquía de RNAs que se encarga de realizar la tarea por etapas.

La jerarquía implementada consiste en un conjunto de Perceptrones Multicapa (PMCs), con un "clasificador modal" que se encarga de clasificar los patrones de entrada de acuerdo a su modo de articulación. Una vez preclasificado, el patrón de entrada es redirigido a un segundo PMC que ha sido entrenado para clasificar los fonemas que componen una categoría modal particular.

Se presentan resultados obtenidos al clasificar datos de 108 locutores del corpus en castellano de OGI, el cual contiene habla continua grabada por línea telefónica. Los resultados expuestos abarcan cuatro parametrizaciones utilizadas comúnmente en el preproceso de señales de voz: CPL(LPC), Cepstrum, PLP y Mel Cepstrum.

1 Introducción

Nuestro grupo de trabajo se encuentra involucrado en el desarrollo de un sistema para el Reconocimiento Automático de Habla Continua y Espontánea (RAHCE). Este sistema se basa en la propuesta conexionista Redes Neuronales Artificiales (RNAs) - Modelos Ocultos de Markov (MOMs), cuyos fundamentos teóricos se establecieron en años recientes [1].

La principal novedad introducida en la propuesta conexionista fue el uso de un Perceptrón Multicapa como estimador probabilístico en vez de los estimadores utilizados comúnmente en los MOMs; frente a éstos el PMC tiene las siguientes ventajas [2]:

1 Los estimadores convencionales de los MOMs se basan en fuertes suposiciones acerca de las características estadísticas de los datos de entrada que son innecesarias en los estimadores basados en PMCs.

2 Los PMCs son una buena alternativa como funciones discriminantes ya que sus parámetros son optimizados para maximizar la discriminación entre clases y no para modelar las distribuciones *per se*.

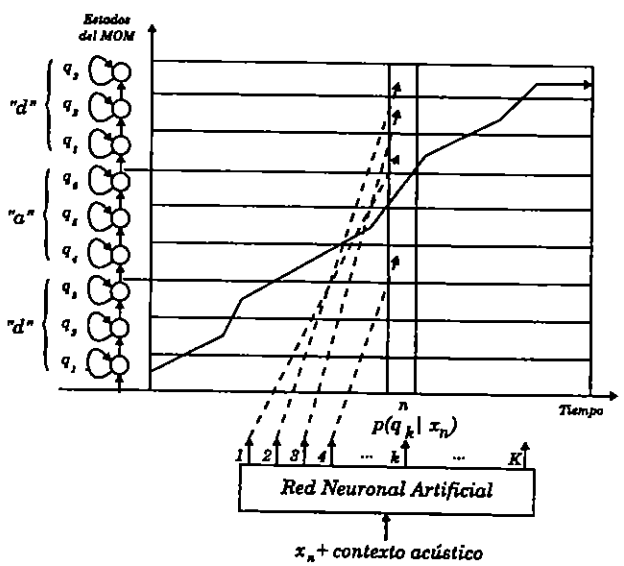


Figura 1: Esquema híbrido MOMs-PMC para la estimación de la probabilidad $p(x_n|qk)$

*Los autores agradecen al Centro para la Comprensión del Lenguaje Hablado (CSLU) de Oregon su amabilidad al proveer el corpus utilizado en este trabajo

Tabla 1: Conjunto reducido de etiquetas (v.g., no diacríticas y agrupadas)

No	ETIQUETA	No	ETIQUETA	No	ETIQUETA	No	ETIQUETA	No	ETIQUETA
1	a	7	p	13	ch	19	rr	25	y
2	e	8	t	14	ll	20	r	26	R
3	i	9	k	15	f	21	m		
4	o	10	b	16	z	22	n		
5	u	11	d	17	s	23	ñ		
6	ai	12	g	18	j	24	l		

En la figura 1 se muestra un esquema del estimador estadístico híbrido MOM/PMC donde a cada paso n el vector acústico x_n y su contexto se presentan como entradas al PMC. Las probabilidades locales se generan por la RNA y son utilizadas, luego de la división por las probabilidades *a priori*, como medidas locales de similitud por el *algoritmo de Viterbi*.

2 Motivación

Ya que la técnica ilustrada en la sección 1 ha sido utilizada durante los últimos años, se decidió utilizarla en el sistema de RAHCE en desarrollo. Para ello se utilizó el corpus de habla telefónica continua de OGI. Este corpus contiene miles de grabaciones multilenguaje que fueron adquiridas en el Instituto de Graduados de Oregon (OGI) por línea telefónica en 22 idiomas [3]. Las principales características del corpus en castellano de OGI son las siguientes:

- Contiene voces de 108 locutores, 74 hombres y 34 mujeres, de habla castellana de todo el mundo.
- Las llamadas fueron adquiridas a 8000 muestras por segundo.
- El contenido por locutor es:
  a) Palabras convenidas y frases: 24 seg.
  b) Descripciones cortas: 42 segundos.
  c) Habla continua inducida: 50 seg.

Algunas partes del corpus están etiquetadas, en particular, los 50 segundos de habla continua inducida; para el etiquetado se utilizaron 204 etiquetas tomadas del alfabeto *Worldbet*. En realidad, hay solamente 55 etiquetas básicas en el corpus, sin embargo, éstas fueron expandidas a 204 con el fin de describir características específicas en algunas emisiones de los fonemas; la expansión se realizó agregando información diacrítica a las 55 etiquetas base.

El conjunto de etiquetas contenidas en el corpus de OGI fue reducido a 26 con el fin de disponer de una transcripción *fonémica* para el castellano. Esta tarea se realizó ignorando la diacrítica y agrupando etiquetas que se consideraron teóricamente equivalentes. El conjunto reducido de etiquetas se muestra en la tabla 1, donde la etiqueta /ai/ ha sido incluida por consistencia con la documentación del corpus [3] y la etiqueta /R/ incluye a todas las etiquetas de *no-habla* (v.g., ruido) definidas en el corpus.

De esta manera, se llevaron a cabo algunos experimentos de clasificación fonética utilizando un PMC de acuerdo a lo mencionado en la sección 1. Con el fin de que la tarea fuera independiente del locutor se separó el corpus en dos grupos de locutores: uno para entrenamiento y otro para prueba. Ambos subconjuntos estaban formados por 37 hombres y 17 mujeres.

La literatura científica sugiere que para realizar una clasificación fonética con un PMC en topología de tres capas, la capa intermedia debe contener entre 500 y 4000 unidades de proceso [2]. Así, se llevaron a cabo experimentos de Clasificación Fonética Dependiente del Locutor (CF-DL) y Clasificación Fonética Independiente del Locutor (CF-IL). Solo se abordará en este documento la CF-IL.

Debe mencionarse que en todos los experimentos reportados en este documento se realizó la parametrización de la señal de voz a una tasa de 10 milisegundos con tramas que cubrían y modelaban 25 ms de señal; esto equivale a decir que cada 10 ms se alimenta el PMC con un patrón de información que representa 25 ms de la señal de voz para entrenamiento o para prueba. El tipo de parámetros empleados en la representación fueron Coeficientes de Predicción Lineal (LPC), el *cepstrum*, coeficientes de Predicción Lineal Predictiva (PLP) y Mel cepstrum (para más información sobre estas técnicas se puede consultar [4]).

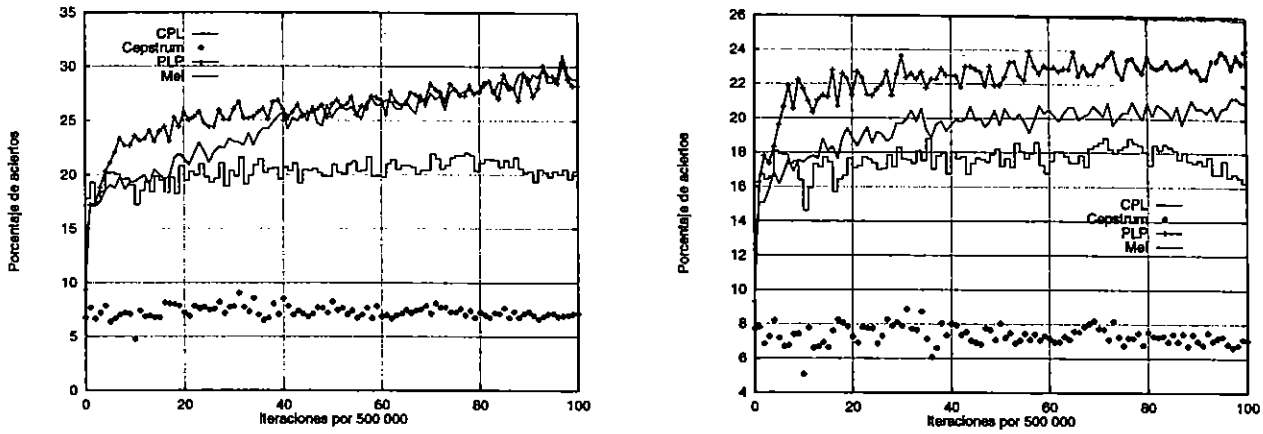


Figura 2: Porcentajes de clasificación correcta en entrenamiento y prueba con un PMC

La capa intermedia en el PMC de CF-DL contenía entre 1000 y 2000 neuronas, mientras que en CF-IL se manejaron únicamente algunos cientos de neuronas por consistencia con experimentos que se mostrarán en la siguiente sección. La figura 2 muestra los resultados obtenidos al clasificar los fonemas de la tabla 1 con un PMC de topología 12x676x26.

El entrenamiento de los PMCs se realizó de manera supervisada dando para cada entrada un patrón de salida deseada; todos los elementos de este patrón son cero excepto el de la neurona asociada a la clase fonética del patrón de entrada, el cual se pone a uno. La eficiencia de la clasificación se midió eligiendo la neurona de la capa de salida con valor más alto y comparando la clase asociada a ella con la clase asociada al patrón de entrada.

Las figuras 2(a) y 2(b) muestran las tasas de acierto al clasificar los fonemas del corpus de entrenamiento y prueba respectivamente. Estas gráficas se han obtenido en función del número de ejemplos utilizados para entrenamiento para los cuatro tipos de parametrizaciones contempladas: Coeficientes de Predicción Lineal (CPL), *Cepstrum*, Predicción Lineal Perceptual (PLP(y *Mel Cepstrum*. La tasa de aprendizaje se fijó a 0.05 y la inercia a 0.8; se actualizaron los pesos cada cien iteraciones.

Analizando las gráficas se observa que si bien se han aplicado $50X10^6$ ejemplos la tasa de aciertos en entrenamiento apenas llega el 30 %, mientras que la de prueba ronda el 24 %. Si bien, el comportamiento asintótico de las curvas de entrenamiento sugiere que puede haber mejoras, las curvas de aciertos en prueba parecen estabilizarse y se puede considerar que aún continuando con el entrenamiento la tasa de aciertos no aumentaría.

Debe subrayarse que los resultados obtenidos no pueden extrapolarse hacia otras técnicas debido a la diferencia en los corpus experimentales utilizados. Se puede acudir a otras fuentes para conocer resultados obtenidos en lenguas diferentes al castellano (v.g., ver [1] para las lenguas inglesa y alemana).

3 La técnica propuesta

El objetivo principal de la técnica propuesta es disponer de un sistema de clasificación fonética más pequeño y eficiente que el sistema monoPMC. Esto es, con un entrenamiento más barato y una tasa de aciertos más alta. La alternativa propuesta consiste en una estructura jerárquica de PMCs en la cual se realiza una preclasificación basada en los modos de articulación del castellano. Las características articulatorias (v.g., modo y punto de articulación) son un medio muy común de análisis y clasificación categórica en la lingüística [5] [6].

En la figura 3 se muestra un esquema de la técnica propuesta. En el entrenamiento, cada PMC se entrena con un conjunto de datos diferentes para un conjunto de clases distintas. Esto es, el PMC modal se entrena con todos los datos de entrenamiento para distinguir los modos de articulación, mientras que los clasificadores fonéticos restantes son entrenados con datos etiquetados con su categoría modal.

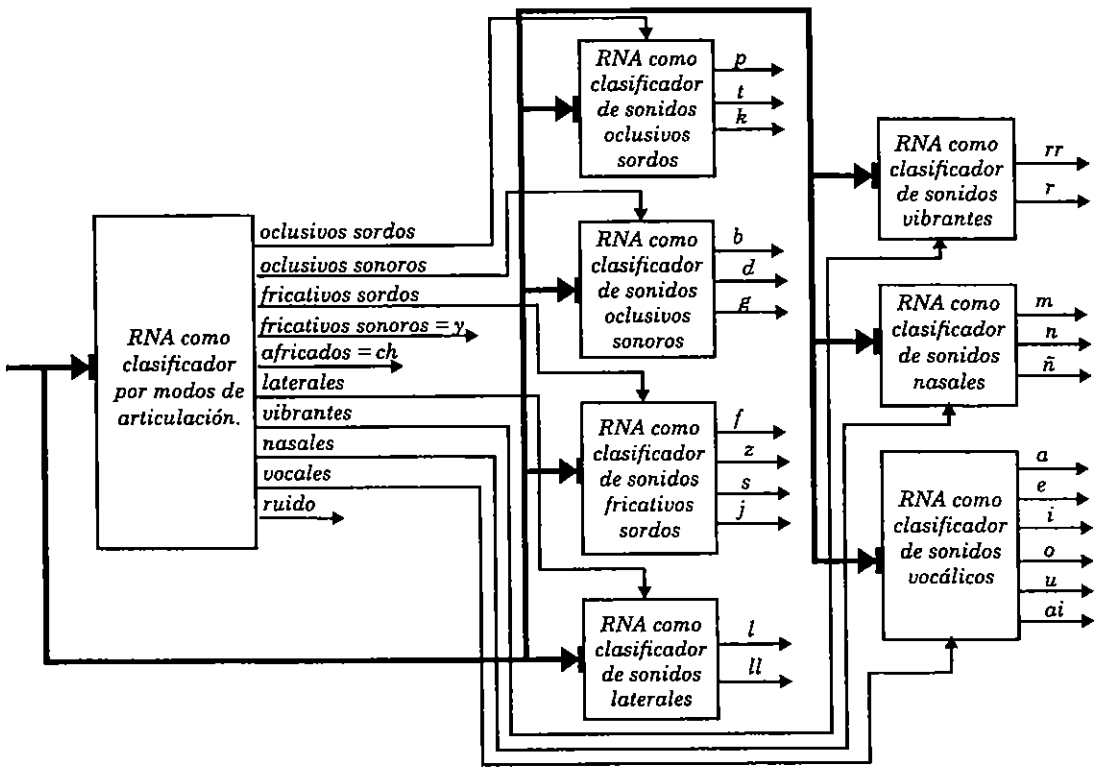


Figura 3: Esquema de la estructura jerárquica para la clasificación fonética

Tabla 2: Equivalencia entre etiquetas

Categoría modal	Fonemas	Worldbet
Oclusivo sordo	p	p pc
	t	t[t[c
	k	k kc
Oclusivo sonoro	b	V b bc
	d	D d[d[c
	g	G g gc
Fricativo sordo	f	f
	z	T
	s	s hs S
	j	x h
Fricativo sonoro	y	j
Africados	ch	tS tSc
Laterales	l	l
	ll	L dZ
Vibrantes	rr	r
	r	r(
Nasales	m	m
	n	n[N
	ñ	nj
Vocales	a	a 3 & @
	e	e E
	i	i I
	o	o 0
	u	u w U
	ai	aI
No habla	R	No incluídas

En la tabla 2 se muestra la correspondencia entre las 26 clases fonéticas, las etiquetas base del corpus y las clases modales. Estas asociaciones se establecieron de acuerdo a Fuentes [6].

Durante la clasificación las salidas del PMC modal indican a cual clasificador han de redirigirse los patrones de entrada para la clasificación final, si bien, éste puede enviarse a todos los PMCs *fonéticos* con el fin de tomar en cuenta todas las salidas de los PMCs en un esquema de RAH mediante alineamiento temporal como se muestra en la figura 1 para el algoritmo de Viterbi. Esta técnica de clasificación está muy relacionada con la Mezcla Jerárquica de Expertos (MJE), otra alternativa reciente ligada al principio *dividir para conquistar* [7].

4 Experimentos y Resultados

Como ya se mencionó anteriormente, los experimentos que se mostrarán fueron realizados con cuatro tipos de parametrizaciones base para la señal de voz: CPL, *Cepstrum*, PLP y Mel *cepstrum*. Se manejaron como parámetros adicionales o extendidos la Energía Normalizada (ENG) y la Tasa de Cruces por Cero (TCC), ambas a nivel de trama, así como información contextual a nivel supratrama.

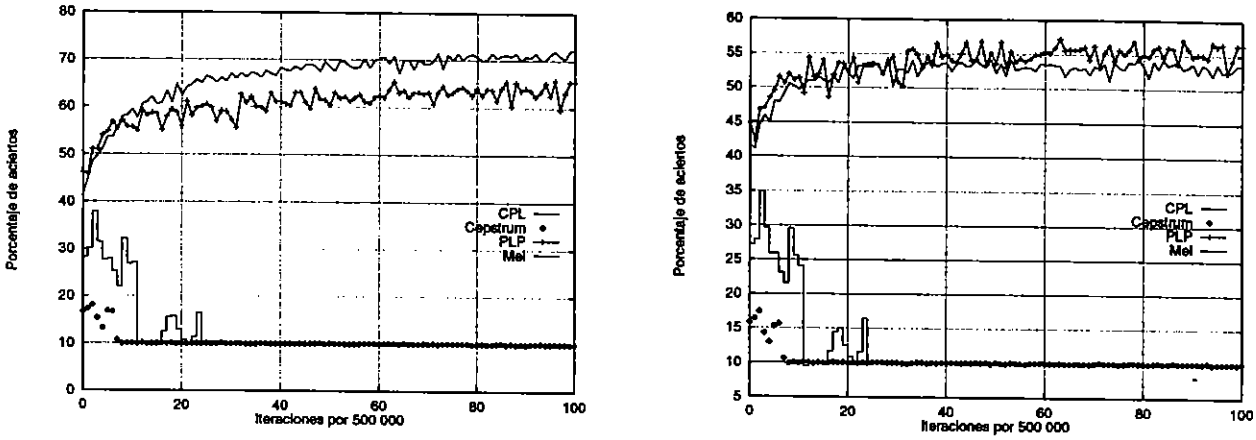


Figura 4: Porcentaje de aciertos en entrenamiento y prueba con el PMC modal

4.1 El clasificador modal

El clasificador modal tiene una topología de 100 unidades en la capa oculta y 10 en la capa de salida; el tamaño de la capa de entrada es variable. El numero de neuronas de la capa intermedia es el cuadrado del número de clases de salida, mientras que la tasa de aprendizaje se fijó en 0.05 y la inercia en 0.8; se actualizaron los pesos cada 100 iteraciones. Estos valores se establecieron de manera empírica mediante experimentos preliminares.

La figura 4 muestra las curvas de aciertos obtenidas con el clasificador modal en una topología 98x100x10 (v.g., tres tramas de contexto a cada lado) como función del número de ejemplos de entrenamiento. La figura 4(a), en particular, muestra los resultados obtenidos al clasificar el corpus de entrenamiento, mientras que la figura 4(b) muestra los resultados para el corpus de prueba.

Puede verse que los resultados obtenidos con CPL y PLP son mejores que los obtenidos con *Cepstrum* y Mel *cepstrum*. Este fenómeno se atribuye a las grandes diferencias en dimensionalidad de los parámetros *cepstrum* y Mel *cepstrum* comparado con CPLs y PLPs. Mientras aquellos tienen valores relativamente altos, éstos son muy cercanos a cero.

Con el fin de disponer de pocos parámetros para evaluar la eficiencia de la tarea, se obtuvieron dos variables de las matrices de confusión generadas por la clasificación: Los Porcentajes Máximo y Promedio de Clasificación Correcta (PMCC) y (PPCC). Éstos se midieron tanto en entrenamiento como en prueba.

Las tablas 3 y 4 muestran los resultados obtenidos con el PMC modal al aplicar patrones *brutos* y normalizados de los cuatro tipos de parametrización utilizados. Una entrada igual a 14 corresponde a 12 parámetros base más Energía (ENG) y Tasa de Cruces por cero (TCC); ambas a nivel de trama. Las siguientes filas corresponden a estos mismos parámetros agregando contextos de 1,2,3 y 4 tramas antes y después de la trama actual.

Los experimentos de la tabla 4 se realizaron con el fin de observar la eficiencia del clasificador cuando la diferencia en la dimensión de los parámetros de entrada es eliminada; esto se realizó normalizando los patrones de entrada al intervalo [-1,1].

Así mismo, ya que los PMCs formados por neuronas con funciones sigmoides tradicionales (v.g., con intervalo de salida [0,1]) invierten gran parte del tiempo inicial de entrenamiento en polarizar sus pesos hacia el valor medio de activación, se cambió la función de transferencia de las neuronas de la capa intermedia a la sigmoide simétrica con intervalo de salida [-1,1] y valor medio 0.

Analizando los resultados de la tabla 4 se observa que al normalizar los patrones de entrada los resultados con CPL decaen ligeramente, comparados con los de la tabla 3. Con los parámetros PLP se obtiene una mejoría muy ligera y con *Cepstrum* y Mel *cepstrum* se producen mejorías importantes. En general, al normalizar las entradas los resultados obtenidos son similares con cualquiera de las cuatro parametrizaciones utilizadas.

Tabla 3: PMCC y PPCC para el clasificador modal con parámetros *brutos*

	CORPUS DE ENTRENAMIENTO							
ENTRADA	CPL		*Cepstrum*		PLP		Mel *cepstrum*	
	PROM	MAX	PROM	MAX	PROM	MAX	PROM	MAX
12	42.92 %	46.85 %	13.55 %	14.69 %	42.42 %	45.84 %	16.81 %	32.43 %
14	43.91 %	47.92 %	13.48 %	16.78 %	43.48 %	46.85 %	17.57 %	34.03 %
42	53.94 %	60.34 %	13.91 %	17.98 %	51.52 %	56.55 %	13.30 %	32.57 %
70	61.41 %	67.82 %	10.76 %	18.39 %	56.30 %	61.59 %	12.11 %	32.85 %
98	66.43 %	72.22 %	10.45 %	18.10 %	60.71 %	65.75 %	12.31 %	37.92 %
126	69.64 %	75.13 %	14.95 %	19.15 %	63.48 %	67.94 %	11.73 %	35.58 %
	CORPUS DE PRUEBA							
ENTRADA	CPL		*Cepstrum*		PLP		Mel *cepstrum*	
	PROM	MAX	PROM	MAX	PROM	MAX	PROM	MAX
12	37.23 %	38.87 %	14.11 %	14.98 %	38.97 %	40.88 %	16.11 %	29.96 %
14	38.25 %	40.70 %	13.93 %	15.17 %	39.81 %	41.60 %	16.95 %	31.68 %
42	43.81 %	46.63 %	14.28 %	16.59 %	45.97 %	49.72 %	12.94 %	30.04 %
70	48.93 %	51.61 %	10.71 %	17.81 %	50.34 %	54.09 %	11.95 %	32.23 %
98	52.22 %	54.38 %	10.38 %	17.48 %	53.87 %	57.21 %	12.04 %	34.99 %
126	54.26 %	56.33 %	14.51 %	18.03 %	55.84 %	58.80 %	11.63 %	33.93 %

4.2 Los clasificadores intramodales

Para cada uno de los clasificadores modales se ha generado un conjunto de gráficas y tablas como las mostradas en la sección anterior para el clasificador modal. Sin embargo, la extensión de este material hace imposible incluirlo en este documento y se opta por mostrar una tabla resumida que muestre la capacidad del CF-IL completo; es decir, clasificador modal y clasificadores intramodales a la par.

La tabla 5 muestra los resultados obtenidos con un CF-IL de 126 entradas (v.g., trama central y cuatro tramas de contexto delante y detrás) utilizando PLPs normalizados con los PMCs de máxima tasa de clasificación; es decir, con el mejor clasificador.

La primera columna muestra las categorías modales y en las columnas segunda y tercera se muestran los resultados de clasificación modal en entrenamiento y prueba respectivamente. La cuarta columna muestra las categorías fonéticas que componen cada categoría modal seguida de dos columnas con los resultados de CF-IL para las clases de ese modo en particular tanto en entrenamiento como en prueba.

Las últimas dos columnas se obtienen al multiplicar los porcentajes modales e intramodales en una relación entrenamiento por entrenamiento y prueba por prueba a nivel de columna; de esta manera se obtienen los porcentajes de CF-IL para cada clase.

En la última fila, al final de la tabla, se muestran los porcentajes promedio obtenidos para cada columna y tarea; esto es: clasificación modal en entrenamiento y prueba, clasificación intramodal en entrenamiento y prueba, y clasificación fonética en entrenamiento y prueba.

Debe notarse que algunas celdas de la tabla se encuentran vacías; en particular, aquellas asociadas a los fonemas /y/ y /ch/ para los cuales no existe clasificador intramodal al corresponder éstos a modos monofonema. Esto es, las salidas del clasificador modal para esas dos categorías corresponden directamente con los fonemas /y/ y /ch/. Tampoco se fue más allá en cuanto a la categoría /ruido/, ya que todos los eventos de *no-habla* se integraron en ella como un todo y no se intentó clasificarlos en subcategorías.

Por otra parte, si bien se mencionó anteriormente que en todos los PMCs se utilizó una topología IxO^2xO, los PMCs intramodales fueron entrenados con una tasa de aprendizaje de 0.2 y se actualizaron los pesos cada 10 ejemplos en vez de los 100 utilizados para el PMC modal; la inercia se mantuvo en 0.8. Esto se hizo debido a la menor cantidad de datos de entrenamiento disponibles.

Tabla 4: PMCC y PPCC para el clasificador modal con parámetros normalizados

	CORPUS DE ENTRENAMIENTO							
ENTRADA	CPL		*Cepstrum*		PLP		Mel *cepstrum*	
	PROM	MAX	PROM	MAX	PROM	MAX	PROM	MAX
12	39.77 %	42.97 %	38.22 %	40.74 %	47.84 %	51.45 %	45.51 %	48.36 %
14	42.14 %	45.69 %	42.99 %	46.10 %	48.24 %	51.41 %	46.83 %	50.50 %
42	51.67 %	57.28 %	55.40 %	61.38 %	60.73 %	66.55 %	59.24 %	64.72 %
70	59.23 %	66.68 %	63.16 %	68.39 %	68.03 %	72.82 %	66.60 %	72.31 %
98	63.68 %	70.72 %	68.40 %	73.16 %	72.83 %	77.52 %	71.56 %	76.78 %
126	67.49 %	72.74 %	71.30 %	75.93 %	75.55 %	79.98 %	73.78 %	79.29 %
	CORPUS DE PRUEBA							
ENTRADA	CPL		*Cepstrum*		PLP		Mel *cepstrum*	
	PROM	MAX	PROM	MAX	PROM	MAX	PROM	MAX
12	34.50 %	36.15 %	35.96 %	37.54 %	42.79 %	44.02 %	41.00 %	42.49 %
14	36.70 %	38.79 %	39.88 %	41.41 %	42.53 %	43.70 %	41.47 %	43.38 %
42	42.13 %	45.61 %	45.97 %	48.07 %	50.40 %	52.62 %	49.62 %	51.95 %
70	46.94 %	49.41 %	50.93 %	52.85 %	54.90 %	56.79 %	54.21 %	56.20 %
98	49.38 %	52.21 %	53.91 %	55.56 %	58.37 %	60.26 %	57.23 %	59.30 %
126	51.88 %	54.90 %	56.23 %	58.28 %	60.24 %	62.29 %	59.54 %	61.07 %

5 Conclusiones

Se ha descrito una técnica de clasificación fonética independiente del locutor utilizando RNAs; esta técnica de clasificación se basa en el principio *dividir para conquistar* al utilizar una estructura jerárquica de PMCs análoga a los modos de articulación del castellano.

El sistema de clasificación descrito en este documento es un sistema integral que analiza, clasifica y proporciona información para todos los fonemas del castellano. Los autores consideran que este tipo de trabajos no abundan y son incluso escasos, en cuanto a divulgación, en otras lenguas de presencia tecnológica más fuerte como el inglés. En ese sentido, este documento constituye una fuente de información.

Si bien el uso de varios PMCs para realizar la CF en la técnica propuesta es claramente más complicado que un sistema monoPMC, conviene observar que en el esquema propuesto, y considerando las capas intermedias y de salida de los PMCs, se necesitan aproximadamente 220 neuronas, mientras que utilizando un PMC se requieren 702. Esto reduce los costos de cómputo y almacenamiento.

En la tarea particular evaluada, CF-IL, los parámetros PLP se mostraron como la mejor opción en detrimento de CPL, *Cepstrum* y Mel *Cepstrum*; sin embargo, el margen de superioridad es muy pequeño cuando se normalizan los patrones de entrada. Este fenómeno se observó también en experimentos de CF-DL llevados a cabo anteriormente por los autores [8].

La técnica propuesta tiene la ventaja de la modularidad, lo cual permite controlar el entrenamiento de cada PMC de manera aislada, además de realizar observaciones locales. Como un ejemplo de ello se puede mencionar que los PMCs más pequeños (v.g., con menos salidas) requieren menos ejemplos de entrenamiento para alcanzar una tasa de clasificación igual o mayor que sus contrapartes más grandes; esto puede comprobarse en la tabla 5.

En algunas categorías modales el uso de PLPs como entrada al clasificador no es la mejor opción, mientras que en otros casos el uso de contexto reduce la eficiencia en la clasificación. Profundizar e ilustrar esta información está lejos del alcance de este documento, aunque se puede decir que este fenómeno sugiere que el desarrollo de un sistema de clasificación fonética con parámetros heterogéneos a la entrada puede producir buenos resultados.

Finalmente, debe mencionarse que la eficiencia observada en el sistema propuesto se degrada al pasar de CF-DL a CF-IL de acuerdo a experimentos llevados a cabo anteriormente [8] y a los datos de la tabla 5. Esto impide que la técnica propuesta se implante directamente en un sistema de RAHC-IL si no se acompaña de algún método de normalización del locutor.

Tabla 5: Resultados de CF-IL para los corpus de entrenamiento y prueba

	% MODO			% INTRAMODO		% NETO	
MODO	**ENT**	**PRUEBA**	**S**	**ENT**	**PRUEBA**	**ENT**	**PRUEBA**
Oclusivo sordo	82.92 %	73.75 %	p	65.08 %	64.67 %	53.96 %	47.69 %
			t	67.39 %	60.12 %	55.88 %	44.33 %
			k	62.20 %	50.77 %	51.58 %	37.44 %
Oclusivo sonoro	61.89 %	50.88 %	b	62.07 %	44.47 %	38.42 %	22.63 %
			d	79.69 %	59.78 %	49.32 %	30.42 %
			g	69.86 %	59.98 %	43.24 %	30.52 %
Fricativo sordo	78.40 %	73.80 %	f	80.01 %	64.15 %	62.73 %	47.34 %
			z	100.00 %	0.00 %	78.40 %	00.00 %
			s	61.37 %	64.56 %	48.11 %	47.65 %
			j	78.92 %	54.22 %	61.87 %	40.01 %
Fricativo sonoro	100.00 %	30.66 %	y			100.00 %	30.66 %
Africados	99.88 %	67.89 %	ch			99.88 %	67.89 %
Laterales	69.76 %	49.72 %	l	80.07 %	78.64 %	55.86 %	39.10 %
			ll	87.51 %	81.20 %	61.05 %	40.37 %
Vibrantes	80.23 %	63.75 %	rr	70.26 %	67.91 %	56.37 %	43.29 %
			r	87.19 %	66.18 %	69.95 %	42.19 %
Nasales	76.82 %	67.40 %	m	73.17 %	65.61 %	56.21 %	44.22 %
			n	60.39 %	55.05 %	46.39 %	37.10 %
			ñ	79.84 %	67.30 %	61.33 %	45.36 %
Vocales	70.12 %	67.03 %	a	70.37 %	64.63 %	49.34 %	43.32 %
			e	71.27 %	73.38 %	49.97 %	49.19 %
			i	73.72 %	72.65 %	51.69 %	48.70 %
			o	65.41 %	65.71 %	45.87 %	44.05 %
			u	82.45 %	68.75 %	57.81 %	46.08 %
			ai	97.72 %	26.36 %	68.52 %	17.67 %
PROMEDIOS	**79.98 %**	**62.29 %**		**75.04 %**	**59.84 %**	**58.95 %**	**39.49 %**

Referencias

[1] Bourlard, H. A. y Morgan N."Connectionist Speech Recognition: A Hybrid approach". Kluwer Academic Publishers, Norwell Massachusetts, USA, 1994, 312 p.

[2] Morgan, N. y Bourlard, H. (1995). "Continuous speech recognition". IEEE Signal Processing Magazine, Mayo de 1995, pp. 25-41.

[3] Lander, T., (1996). "The CSLU Labeling Guide". Internal Report. Center for Spoken Language Understanding (CSLU) of the Oregon Graduate Institute (OGI), Beaverton, Oregon, USA, 93 p.

[4] Picone J. W., 1991. "Signal Modeling techniques in speech recognition". Proceedings of the IEEE, Vol. 79, No. 4, April 1991.

[5] Quilis, A. "Fonética acústica de la lengua española". Serie Biblioteca Románica Hispánica. Editorial Gredos. Madrid, España, 1988, 502 p.

[6] Fuentes, J. L. "Gramática moderna de la lengua española" Serie biblioteca didáctica. Editorial Universitaria, Santiago de Chile, 1991, 520 p.

[7] Jordan, M. I. y Jacobs, R. A., (1993). "Hierarchical mixture of experts and the EM algorithm". Technical Report 9301, Department of Brain and Cognitive Sciences (MIT), U.S.A., 34 p.

[8] Silva-Varela H. y Cardeñoso-Payo V. "Phonetic Classification in Spanish Using a Hierarchy of Specialized ANNs". Proceedings of the 6th Ibero-American Conference on AI: IBERAMIA 98. Lisbon, Portugal, October 1998, pp. 373-384.

Así como en la enseñanza presencial el coste de la plantilla está ligado al numero de alumnos, en el modelo industrial de educación a distancia éste es relativamente independiente de ese factor. En enseñanza a distancia hay un coste fijo, ligado a la preparación de material, mientras que el número de profesores por alumno soporta un ratio mucho más elevado que en la enseñanza presencial.

Una mejora evidente que aporta el uso de las nuevas tecnologías en el modelo industrial es la distribución de material didáctico en soporte informático, ya sea a través de CD-ROM, o de Internet. Utilizar el ordenador como soporte permite integrar presentaciones con simulaciones, herramientas interactivas de solución de problemas, laboratorios virtuales, etc. para promover en los estudiantes un proceso activo de aprendizaje. Más problemática es la introducción de tecnologías de interacción persona-persona. Hay que tener muy en cuenta el no crear una oferta potencial de carácter personalizado que no se pueda luego satisfacer por falta de recursos humanos. Una idea a explorar, de indudable interés, es aumentar el contacto y la colaboración entre estudiantes, aunque también es conocido que es necesario fomentar un cambio de cultura para que la colaboración no solo sea técnicamente posible sino socialmente realizable.

El campus virtual [1] es actualmente una metáfora muy evocada. Se presenta algunas veces como un puente para que las universidades tradicionales extiendan sus actividades, y lleguen a los estudiantes que no se dedican a estudiar a tiempo completo, otras veces se propone como una oportunidad para que las Instituciones de Enseñanza a Distancia incorporen los métodos clásicos de enseñanza, las clases magistrales, junto con el material didáctico para dar mayor soporte al estudio individual a distancia. El problema de las metáforas es que hay que analizar cuidadosamente como se trasladan y aplican. Por ejemplo, a pesar de la corta historia de la videoconferencia multipunto , hay experiencias que manifiestan el fracaso de la misma para dar clases convencionales. El coste del equipo, inasequible al estudiante medio, es una razón evidente, pero otros aspectos, de carácter práctico y organizativo, como por ejemplo que el típico estudiante a distancia no acepta horarios regulares, y tiene unas restricciones fuertes para participar en eventos planificados para grupos, no se han tenido en cuenta al proyectar la disponibilidad de un estudiante convencional al caso de una situación a distancia.

La tecnología ofrece muchas posibilidades, pero hay que tener muy en cuenta en qué entorno y para qué situación de enseñanza puede realmente aportar un valor añadido. Solo así será realmente aceptada por los usuarios.

En la última década la Comunidad Europea ha lanzado una variedad de programas para promover la cooperación internacional en educación. Algunos programas se dedicaron a estudiar la situación en los diferentes ámbitos y niveles de formación, para recomendar acciones futuras. Otros se centraban en el desarrollo de proyectos de I+D, más orientados hacia la tecnología, para explorar las posibilidades de un aprendizaje flexible y a distancia. También se han llevado a cabo un conjunto de experiencias piloto, así como diversas aplicaciones en dominios muy variados..

En el marco del programa SOCRATES [2], ACO*HUM [3] es una amplia red temática en la que participan numerosas Universidades. Esta red incluye un grupo de trabajo especializado en Lingüística Computacional e Ingeniería del Lenguaje. En cooperación con ELSNET [4], se propuso un plan para desarrollar 6 cursos piloto para educación a distancia. Nosotros propusimos uno sobre el tema de Técnicas de PLN para Recuperación de Información [5]. El proyecto fue aprobado en Febrero del 98, y ha sido desarrollado a lo largo de dicho año.

Este tema es especialmente apropiado para abordarlo mediante el uso de la web, con un enfoque de aprendizaje activo por parte del estudiante. Internet ofrece no solo un inmejorable campo de experimentación sino que además los buscadores son las aplicaciones más relevantes de las técnicas de recuperación de información. Se puede guiar a los estudiantes para que utilizando software de PLN disponible on-line, puedan tratar,

expandir, traducir, .. preguntas en lenguaje natural de forma que tengan una experiencia directa del interés y la utilidad de dicho tratamiento para mejorar los resultados de la búsqueda.

En la sección siguiente describimos el enfoque con el que hemos construido el curso. En la sección tercera se ofrece una descripción más detallada de la estructura y contenido del mismo.

2 *Enfoque y objetivos*

Hemos partido por un lado de la experiencia en elaborar material didáctico para el estudio individual a distancia, y por otro hemos tenido muy en cuenta dos principios que se postulan en las teorías recientes sobre el aprendizaje, en particular los conceptos de (i) *aprendizaje activo* y (ii) *aprendizaje situado*. El primero [6] incide en los procesos de adquisición del conocimiento, mientras que el segundo [7] se preocupa sobre la forma en que dicho conocimiento se contextualiza para poder aplicarlo en la vida real.

Un primer aspecto de nuestro enfoque es ofrecer un acceso estructurado a un conjunto de herramientas y recursos. Algunos de ellos se encuentran en nuestro servidor, y otros son externos. Para facilitar que los estudiantes aprendan haciendo, se les pide que realicen para cada idea clave, un conjunto de tareas. Estas tareas se han diseñado como trabajos prácticos, que sitúan la experiencia de aprender en un entorno auténtico, cercano al desarrollo de un caso real.

Un segundo aspecto es ayudar al estudiante a conocer y participar en una comunidad profesional. A este respecto nos proponemos establecer vías de colaboración con todos los posibles interesados: investigadores, profesores, profesionales y estudiantes, para mantener nuestro producto al día. Este punto es importante en este tema que evoluciona tan rápidamente

Un tercer aspecto es ofrecer al estudiante aislado la oportunidad de formar parte de un grupo virtual. Un entorno que soporte a grupos, con objetivos comunes y facilitando oportunidades de participación, son elementos que mejoran en los estudiantes la motivación de aprender. Por ello se incluyen facilidades para utilizar espacios virtuales compartidos, además de medios personalizados de comunicación electrónica.

Finalmente, el último aspecto está relacionado con el papel de la colaboración para el aprendizaje a distancia. Como hemos mencionado previamente, se necesita desarrollar una cultura de colaboración, de cara al estudio y a la futura vida profesional. Con el propósito de explorar el aprendizaje en grupo, incluiremos en nuestro "site" la posibilidad de organizar pequeños grupos de estudiantes que desarrollen colaborativamente actividades, utilizando tecnología asíncrona..

Teniendo en cuenta estos aspectos, los objetivos que nos hemos planteado en el proyecto son los siguientes:

- Desarrollar un curso
  - Que trate el tema de la utilización de técnicas de procesamiento de lenguaje natural para recuperación de información, o para ser más exactos, de recuperación de textos, planteando los problemas de multilingualidad en las fuentes o en la expresión de las preguntas.
  - Que incluya un conjunto de recursos y herramientas on-line, para experimentar las ideas y conceptos que se tratan en el curso. Algunas herramientas software se instalarán en nuestro servidor, como por ejemplo: un stemmer, un analizador morfológico, un tagger, o una base de datos léxica multilingüe , otras son de dominio público, como buscadores o sistemas de traducción, y están disponibles en la red.
- Diseñar un "site" web
  - Que permita aprendizaje individual on-line, y en modo colaborativo
  - Que ofrezca ayuda y soporte a los estudiantes aislados
  - Que disponga de una interfaz que facilite un acceso flexible tanto al contenido del curso, como a los

recursos y a las herramientas disponibles.

- Involucrar a colegas de otros centros para que contribuyan y participen en el desarrollo futuro del prototipo.

En las secciones restantes describimos cómo hemos plasmado estas ideas de forma concreta en el diseño del curso, las pruebas a realizar con usuarios, así como la evaluación interna del mismo y los planes para su diseminación.

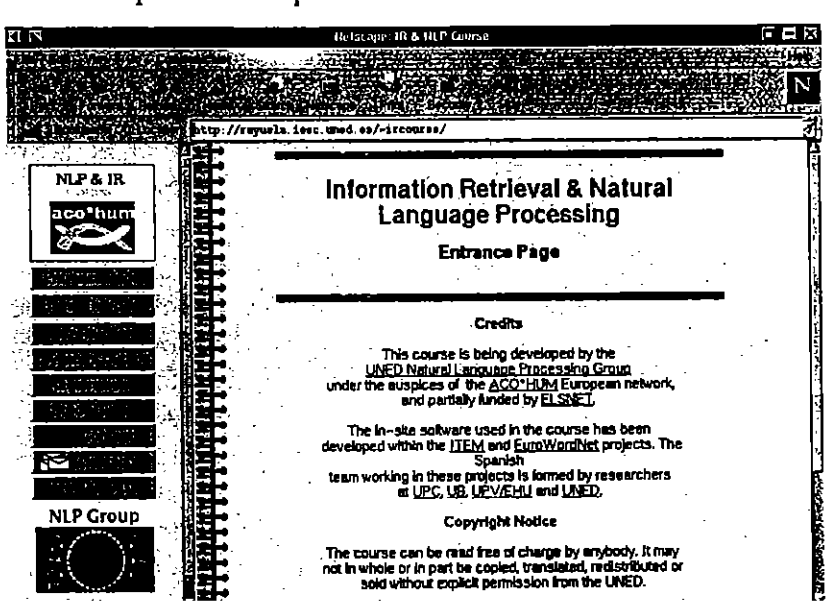
3 *Características principales del IR-NLP web site.*

La figura 1 muestra la página principal de nuestro Web Site. La plantilla es homogénea para todas las páginas: a la izquierda aparece el conjunto de opciones, a la derecha el contenido de la función seleccionada. Hay dos menús diferentes: uno para la página principal, otro para el contenido del curso. El menú de la página principal (ver figura 1) ofrece información general y los servicios de comunicación, además del acceso al contenido del curso. La información general comprende cinco secciones: *Using this site*, *Introduction*, *Syllabus outline*, *Requirements* and *Study Guide*. Por ejemplo *Introduction* responde a las siguientes preguntas:(i) ¿ De qué trata este curso ? (ii) ¿ Qué es lo que no se encontrará aquí ? (iii) ¿ Cómo se puede utilizar este site para aprender? Los servicios de comunicación incluyen correo electrónico con una lista de contactos, y unas *News*. También se ha previsto una sección de preguntas frecuentes, para dar respuesta a los problemas que se planteen. Esta sección se construirá y enriquecerá con las contribuciones de los usuarios

El botón *Contents* da acceso a la página principal del curso. El menú de esta página (ver figura 2) contiene las siguientes funciones :

Contents – muestra el Indice del curso. El curso está organizado en cuatro capítulos, cada uno cubre un tema, y se estructura en un conjunto de secciones. El botón muestra la lista de secciones del capítulo en curso. Se puede navegar por las páginas correspondientes, pinchando en el subtitulo del índice. Al principio se ven los títulos de los cuatro capítulos: *Overview*, *Information Retrieval*, *NLP in IR* and *Cross-Language IR*. En la parte derecha aparece el contenido de cada sección, en donde se presenta la materia, junto con ejemplos y ejercicios interactivos. La presentación es fundamentalmente hiper-textual, con enlaces a una lista de lecturas y al glosario. Se incluye también en cada capítulo una prueba de auto-comprobación, que puede rellenarse interactivamente y enviarse, para una evaluación automática. Los ejercicios tienen una descripción didáctica, que les caracteriza en los siguientes términos:

- *Estimated workload*: Da una estimación del tiempo necesario para resolver el ejercicio.
- *Difficulty level*: se consideran tres niveles, facil, medio y dificil
- *Learning Objectives*: una descripción del objetivo principal del ejercicio



- *Character*: Recomendado u Opcional.
- *Media needed*: lapiz y papel, software interno, o enlaces a fuentes externas.
- *Solution*: Hay tres posibilidades: (i) Se da la solución (ii) Se facilita un método para comprobar la solución (iii) Se proponen preguntas para pensar acerca del resultado.

La Figura 2 muestra uno de los ejercicios del capítulo de Recuperación de Información

Study Guide – ofrece soporte didáctico para llevar a cabo un estudio individual del contenido del curso. Para cada capítulo, se incluyen los siguientes elementos: (i) Una lista de los conceptos, principios o técnicas fundamentales. (ii) Objetivos de aprendizaje, una descripción de los conocimientos o habilidades que el estudiante adiquirirá al final de su proceso de aprendizaje. (iii) Planificación, en dónde se sugiere un orden de estudio y realización de actividades. La guia de estudio facilita un tour guiado del curso. Es el camino que se recomienda para el estudiante independiente que quiera realizar el curso completo.

El resto de los botones dan acceso directo a las colecciones on-line de referencias *References*, enlaces a "sites" externos *Links*, el glosario *Glossary*, los ejemplos *Examples*, los ejercicios *Exercises*, pruebas de auto evaluación *Self-tests* y herramientas software *Software*. De esta forma nuestro site puede ser utilizado como una fuente complementaria bien por profesores, estudiantes o profesionales que deseen enriquecer su propio entorno y recursos de aprendizaje

Por ejemplo con el botón de enlaces ***Links*** se accede a una página en donde todos los enlaces externos se listan organizados bajo los siguientes epígrafes:

- On-line Search Software
  - Search Engines
  - Cross-Language Text Retrieval Search Engines Demos
- Language Resources, Tools and Services
  - On-line dictionaries
  - Lexical-Semantic Knowledge Bases
  - Morphological analyzers and taggers
  - Text Summarization and Information Access

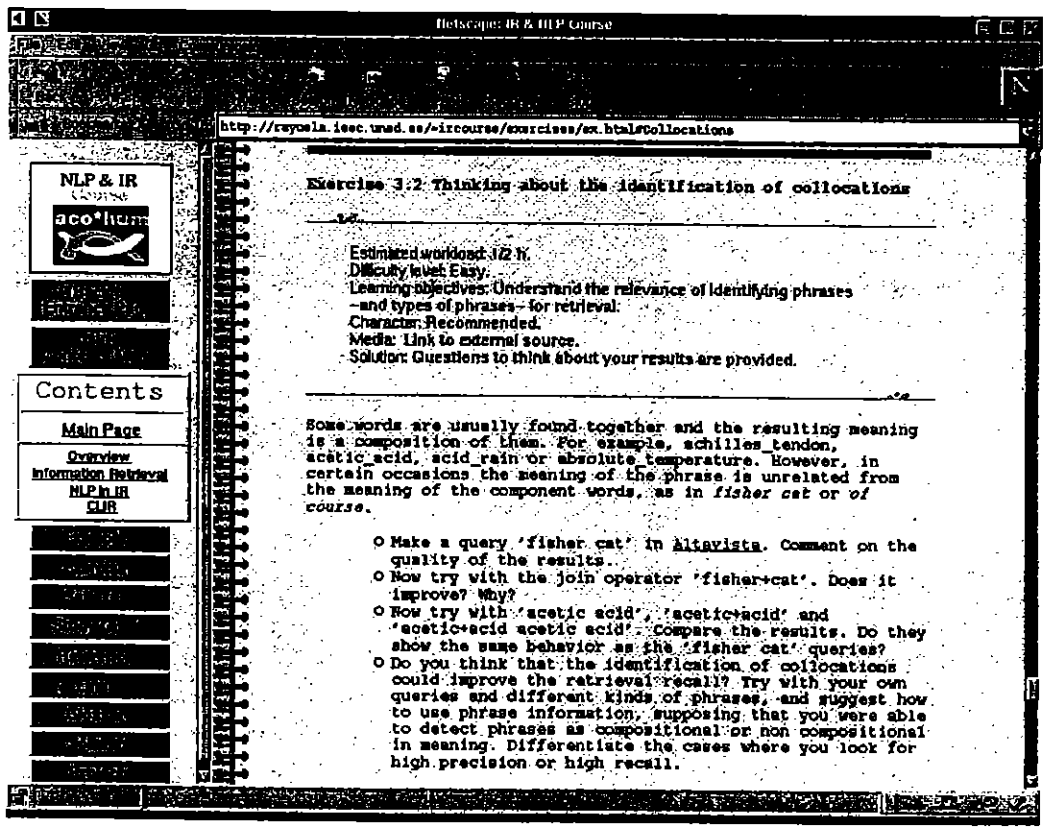


***Figure 2**: Exercise 2.0. Introduction to IR Engines.*

- Machine Translation Services
- Sites of interest
  - Information Retrieval
  - Information Retrieval and Natural Language Processing
  - Cross Language Text Retrieval

References – Esta es la principal fuente documental para seguir el curso. Hay tres tipos de referencias: material procedente de tutoriales o libros, artículos relevantes, y revisiones o síntesis sobre el estado del arte y perspectivas de futuro. Hemos seleccionado

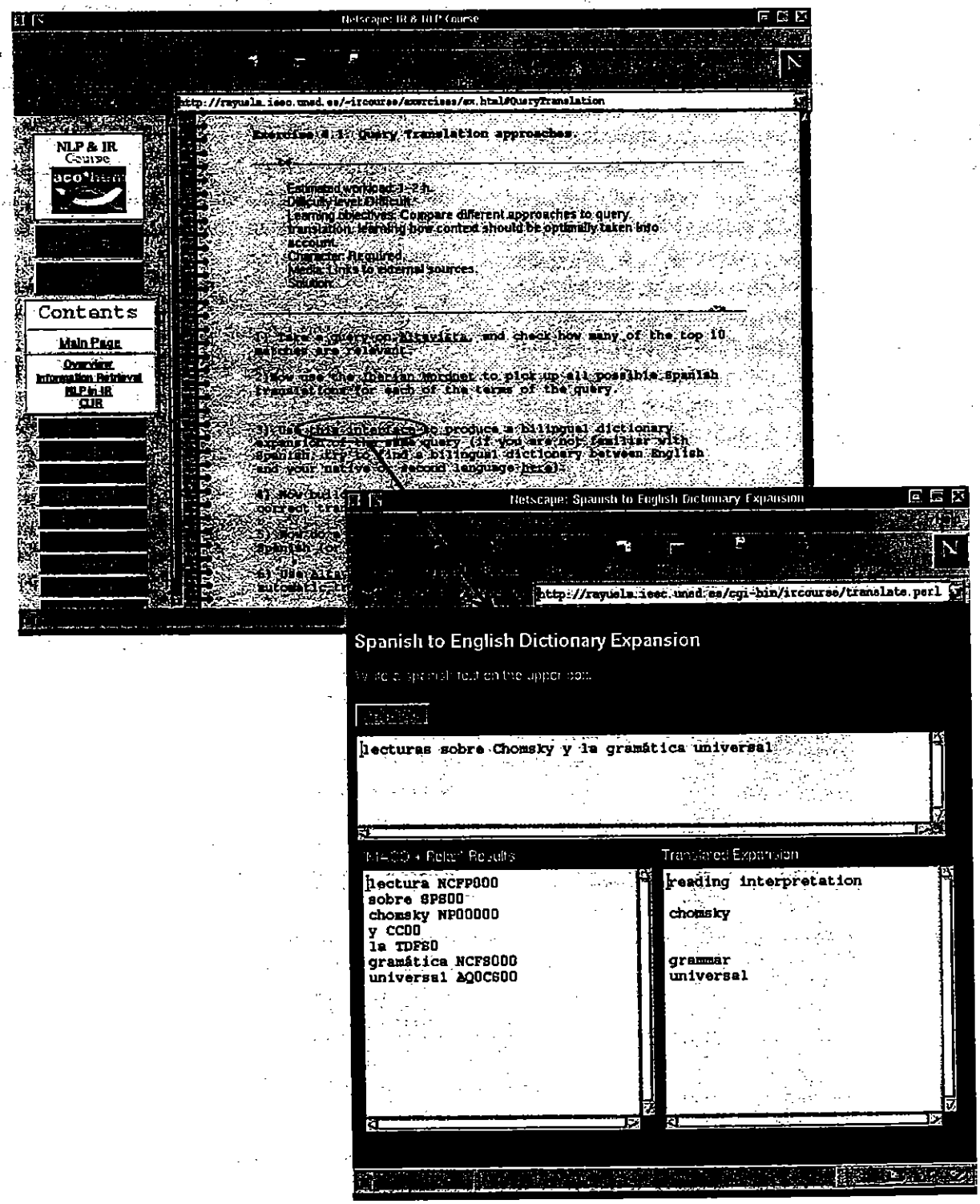


Figura 3. *Ejemplo de interacción con el software*

unicamente material de libre disposición para que el alumno a distancia no tenga que asumir la tarea suplementaria, y a veces costosa, de obtener las referencias a través de centros de documentación.
El botón de ***Software*** lleva a una página donde se describen un conjunto de herramientas relacionadas con la temática del curso. En algunos casos se pueden usar interactivamente, en otros se pueden descargar e instalar en el ordenador propio. Las herramientas son de tres tipos: de dominio público, adaptadas para el curso (como por ejemplo el algoritmo de Porter) o con licencia especial para ser utilizadas en el curso. Para algunas de estas últimas, hemos diseñado y construido interfaces web que facilitan su uso interactivo. La Figura 3 muestra un ejercicio que hace uso de enlaces externos (en las partes 1, 2 y 6) y de recursos internos (parte 3). Pulsando en "*this interface*", aparece una ventana en donde se puede escribir un texto y obtener su traducción palabra a palabra . Los resultados del análisis morfológico y del tagger pueden verse en la parte izquierda. Para las palabras categorizadas en estos procesos de análisis, la traducción basada en consulta a un diccionario, aparece en la ventana derecha.

El botón ***Projects*** es para grupos de estudiantes que sigan estudios reglados, bajo responsabilidad de un profesor, que se encarga de su monitorización. Da acceso a un entorno que facilita la realizacion de actividades colaborativas.

4 Estado actual y desarrollo futuro

El curso se encuentra actualmente accesible en la web, y es de uso público. Aún no consideramos que esté terminado. Nuestra idea es enriquecer y mejorar este prototipo con la colaboración de otros especialistas, así como con la participación de profesores y estudiantes. En una primera fase nos dirigiremos a un pequeño grupo de expertos, y al mismo tiempo probaremos el curso con nuestros estudiantes de doctorado. Más tarde realizaremos una campaña de difusión a través de redes y grupos de interés, así como en congresos del área.

A los expertos les vamos a pedir su opinión sobre los contenidos y su disponibilidad para colaborar ofreciendoles un abanico de posibilidades. Desde pedir que se nos notifique los cambios que se produzcan en sus URL para poder mantener actualizada nuestras listas de enlace, hasta la posibilidad de que contesten preguntas para crear incrementalmente unas FAQ, o ser incluidos en las listas de contacto, para que los estudiantes puedan establecer un contacto directo.

Con los estudiantes nos proponemos evaluar tres aspectos: usabilidad del entorno, interés y adecuación del contenido, y contrastación de las estimaciones que se dan en la guía con los datos obtenidos de la observación de la utilización del curso en condiciones reales.

En la medida en que seamos capaces de establecer y mantener un marco cooperativo, recogiendo y ofreciendo información relevante y actualizada, nuestro sitio web será una referencia útil para los interesados en el área.

Agradecimientos

Queremos agradecer la labor del Dr. K. de Smedt, en la puesta en marcha de la iniciativa ELSNET LE Training Showcase así como a ELSNET por la financiación parcial del proyecto.

Referencias

[1] Verdejo, F., and Davies, G. Editors (1998). *The Virtual Campus: Trends for Higher Education and Training*. Chapman & Hall.
[2] SOCRATES home page:
http://europa.eu.int/en/comm/dg22/socrates.html
[3] ACO*HUM home page:
http://www.hd.uib.no/AcoHum
[4] ELSNET home page: http://www.elsnet.org
[5] IR & NLP Course home page:
http://rayuela.ieec.uned.es/~ircourse
[6] Schank, R. C. (1994). Active Learning through multimedia. *IEEE Multimedia*. Vol. 1(1), Spring 94, pp 69-78.

[7] Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press.