

## ALGUNOS ASPECTOS DEL EUSKARA UTILIZADO HOY

M.<sup>a</sup> José AZURMENDI AYERBE

(En estos últimos años me he dedicado al estudio del «euskara utilizado actualmente» por parte de nuestra población [en Euskadi Sur], motivada por la constante y rápida transformación con que se nos manifiesta hoy el euskara, al menos al nivel de su «uso» que es lo que a mí me ha interesado más. Esta transformación puede ser debida a razones de distinto tipo: el contacto creciente con el español [nuestros vascófonos son en su mayoría, y cada vez más, bilingües], el proceso de normativización y normalización del euskara, la renovación vertiginosa de las cosas y conceptos en nuestro mundo moderno y que lleva a una renovación léxica también vertiginosa, etc. Como homenaje a nuestro gran estudioso del euskara Pierre LAFITTE, quiero colaborar al conocimiento mayor de este euskara nuestro.)

.  
.  
.

En este trabajo vamos a ofrecer algunas de las conclusiones a las que llegábamos en mi Tesis de Doctorado: *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la Comarca de San Sebastián*, defendida en la Universidad Complutense de Madrid en Junio de 1982.

Se trata de algunas de las conclusiones, desde el punto de vista más cuantitativo, obtenidas en el estudio del «léxico disponible» producido por los escolares de 5.<sup>o</sup> y 8.<sup>o</sup> de EGB de las Ikastolas de la Comarca de San Sebastián (Donostia, Pasaia, Lezo, Errenteria, Oiartzun, Hernani, Urnieta, Lasarte, Usurbil), en relación a estos cuatro Centros de Interés (C.I.) o mundos significativos ampliados:

- C.I. 1: Gorputzaren atalak / Partes del cuerpo humano.
- C.I. 2: Jantziak eta oinetakoak / Prendas de vestir y calzado.
- C.I. 10: Jolasak. Atsedendialiak. Desborapasak / Juegos. Diversiones. Entretenimientos.
- C.I. 14: Nekazaritza eta landareak. Abeltzaintza eta abereak. Arrantza eta arrainak / Agricultura y plantas. Ganadería y animales. Pesca y peces.

El estudio del léxico disponible está basado en el análisis del propio léxico, a través de un conjunto de diccionarios, y en el análisis de un conjunto de aspectos o variables lingüísticas relacionadas con dicha producción léxica. Las conclusiones de las que hablaremos en este trabajo están relacionadas con el «uso lingüístico» del léxico, estudiado a través de las variables lingüísticas siguientes:

- Tipo de variables lingüísticas 1, de la cantidad de producción léxica, a través de la variable lingüística. *IPE* = total de ocurrencias/persona.

— Tipo de variables lingüísticas 2, del grado de lexicalización o de la clase léxica producida, a través de las variables:

- 2NO = proporción de sustantivos o nombres producidos;
- 2VE = proporción de verbos producidos;
- 2AJ = proporción de adjetivos producidos;
- 2AD = proporción de adverbios producidos;
- 2VN = proporción de expresiones o perífrasis de verbo + nombre producidas;
- 2CS = proporción de nombres utilizados en un caso determinado, o declinados en euskara, o a través de la preposición en español;
- 2PV = proporción de expresiones o perífrasis de preposición + verbo (producidas sólo en español: «a coger»);
- 2OR = proporción de expresiones que formalmente constituyen una oración («a la una salta la mula», como expresión del nombre de un juego).

— Tipo de variables lingüísticas 3, de la distancia interlingüística, a través de las variables:

- 3DI = proporción de expresiones y palabras fonológicamente diferentes en las dos lenguas en contacto (el euskara y el español). Por ejemplo, entre «burua» y «cabeza»;
- 3PA = proporción de expresiones y de palabras fonológicamente parecidas en las dos lenguas. Por ej., entre «galtzetina» y «calcetín»;
- 3SU = proporción de expresiones y de palabras-sustituto, es decir, los casos de sustitución lingüística. Por ej., en el euskara: «pantaltoa» en vez de «galtza», para «pantalón» en español;
- 3NP = proporción de nombres propios, generalmente préstamos recientes en el euskara desde el español, sobre todo en los C.I. 2 (Prendas de vestir y calzado) y 10 (Juegos, diversiones, entretenimientos). Por ej.: «Adidas», «Superman», etc.

— Tipo de variables lingüísticas 4, de la corrección o grado de normativización utilizada, a través de las variables:

- 4BI = proporción de expresiones y de palabras bien o correctamente utilizadas;
- 4MA = proporción de expresiones y de palabras mal o incorrectamente utilizadas.

— Y tipo de variables lingüísticas 5, de las divisiones significativas o semánticas distinguidas en cada C.I., a través de las variables:

— En el C.I. 2 (Prendas de vestir y calzado):

- 5VS = proporción de expresiones y de palabras que significan respecto del vestido;
- 5CZ = proporción de expresiones y de palabras que significan respecto del calzado;
- 5RE = proporción de expresiones y de palabras de difícil clasificación significativa, el resto significativo.

— En el C.I. 10 (Juegos. Diversiones. Entretenimientos):

- 5CA = proporción de expresiones o de palabras que significan juegos, diversiones o entretenimientos a realizar sobre todo en casa;
- 5CL = proporción de expresiones y de palabras que significan juegos, diversiones o entretenimientos a realizar sobre todo en la calle;
- 5CC = proporción de expresiones y de palabras que significan juegos, diversiones o entretenimientos a realizar parecidamente en casa y en la calle;
- 5DP = proporción de expresiones y de palabras que significan deportes utilizados como juegos-diversiones-entretenimientos.
- 5RE = (resto significativo).

— En el C.I. 14 (Agricultura y plantas. Ganadería y animales. Pesca y peces):

- 5AG = proporción de expresiones y de palabras que significan respecto de la agricultura y plantas;

5GN = proporción de expresiones y de palabras que significan respecto de la ganadería y animales;  
 5PS = proporción de expresiones y de palabras que significan respecto de la pesca y peces;  
 5RE = (resto significativo).

I

Desde el punto de vista de la producción de las distintas variables lingüísticas estudiadas, y de la dispersión con que se ha dado dicha producción lingüística, ofrecemos el siguiente cuadro general:

		$\bar{x}$ - %				$\sigma$			
		C.I. 1	C.I. 2	C.I. 10	C.I. 14	C.I. 1	C.I. 2	C.I. 10	C.I. 14
$\bar{x}$	1PE	0,0347	0,0363	0,0419	0,0311	1,05	0,82	1,49	0,995
	OC/PE	28,82	27,55	23,87	32,15				
%	2NO	94,50	95,45	66,99	97,59	7,32	3,61	8,37	1,98
	2VE	5,49	4,34	12,71	2,21	7,32	3,56	6,14	1,79
	2AJ			0,17				0,42	
	2VN			8,66				5,05	
	2CS			9,62				5,94	
	2OR			0,55				0,44	
	2PV			0,73				1,01	
	2AD			0,50				0,97	
	2RE						0,15		
	3DI	91,99	66,70	66,94	83,13	7,89	34,36	28,51	16,50
	3PA	4,73	28,55	25,35	11,47	5,44	30,66	26,39	12,87
	3SU	3,26	3,80	3,75	5,39	2,77	3,87	3,78	3,87
	3NP		0,92	3,85			0,72	2,29	
	4BI	82,01	78,50	71,00	82,04	10,54	7,98	11,95	9,38
	4MA	17,95	21,49	28,99	17,95	10,56	7,97	11,95	9,37
	5CA			28,25				4,30	
	5CL			26,87				4,27	
	5CC			13,44				2,22	
	5DP			24,78				5,22	
	5AG				26,47				6,79
5GN				47,93				8,78	
5PS				24,46				3,68	
5VS		77,10				2,46			
5CZ		22,37				2,34			
5RE		0,51	6,02	1,11		0,97	3,97	0,86	

En este cuadro pueden verse tanto la producción lingüística, a través de las  $\bar{x}$  = medias, y de las % = proporciones, como la dispersión en que se ha dado dicha producción lingüística, a través de la  $\sigma$  = desviación típica, respecto de cada una de las variables lingüísticas distinguidas dentro de cada tipo de variables, y en relación a cada uno de los cuatro C.I. estudiados.

1. Respecto del tipo de variables lingüísticas 1, de la cantidad de producción léxica:

— El mundo significativo o C.I. en que se ha producido mayor cantidad media de léxico ha sido sobre todo el C.I. 14 (Agricultura y plantas...), en parte explicable por concentrar en uno sólo lo que tradicionalmente ha solido estudiarse como tres mundos significativos distintos: el de la agricultura, el de la ganadería y el de la pesca. Y después el C.I. 1 (Partes del cuerpo humano). Es decir, los mundos disponibles significativamente más tradicionales.

— El mundo significativo o C.I. en que se ha producido una menor cantidad media de léxico ha sido sobre todo el C.I. 10 (Juegos...). Y después, también con bastante diferencia, el C.I. 2 (Prendas de vestir...). Es decir, los mundos disponibles significativamente más innovadores.

2. Respecto del tipo de variables lingüísticas 2, del grado de lexicalización o de la clase léxica producida. Habiéndoseles pedido explícitamente sustantivos y verbos, lo realmente obtenido en las encuestas ha sido:

— Sustantivos o nombres, y verbos, con un resto insignificante, en los mundos significativos o C.I. 1 (Partes del cuerpo), 2 (Prendas de vestir y calzado) y 14 (Agricultura y plantas. Ganadería y animales. Pesca y peces).

— Además de sustantivos y verbos, otras expresiones léxicas distintas en el C.I. 10 (Juegos. Diversiones. Entretenimientos). Entre estas otras expresiones, destacan cuantitativamente: 2CS (proporción de nombres utilizados en un caso determinado, a través de la declinación en euskara y de la preposición en español; por ej.: «izkutaketan» = «al escondite»), y 2VN (proporción de expresiones o perífrasis de verbo + nombre; por ej.: «musika entzun» = «oír música»).

Es decir, se trata de expresiones y perífrasis léxicas y no de palabras simples, por lo que a veces hablamos de grados de lexicalización, y otras veces de lexicalización deficiente.

También desde este punto de vista se nos muestra el mundo significativo de este C.I. 10 (Juegos...) como el menos establecido, el que está en una situación de mayor evolución y cambio.

— Hay un claro predominio de sustantivos o nombres en los cuatro C.I., y sobre todo en el C.I. 14 (Agricultura...).

— Si tenemos en cuenta que todas las otras expresiones léxicas formalmente obtenidas fuera de los sustantivos y verbos, están de algún modo ligadas al verbo, ya que gramaticalmente contienen o suponen un verbo subyacente, obtenemos que el mundo significativo del C.I. (Juegos...) se nos muestra muy diferente: porque frente a un 67% de nombres, hay hasta un 33% de expresiones o palabras ligadas al verbo. Es como si este mundo significativo del juego fuese pensado y expresado de un modo más dinámico.

3. Respecto del tipo de variables lingüísticas 3, de la distancia interlingüística, o del grado de penetración lingüística, entre el léxico obtenido en nuestras dos lenguas en contacto, el euskara y el español:

— El mundo significativo del C.I. 1 (Partes del cuerpo) sobre todo, y también el del C.I. 14 (Agricultura...), los más tradicionales significativamente, son también en los que se mantiene una distancia interlingüística mayor, es decir, donde se aprecia una menor influencia de una lengua sobre la otra, que en nuestro caso lo es del español sobre el euskara.

— Por el contrario, vuelven a ser los mundos significativos de los C. I. 10 (Juegos...) y 2 (Prendas de vestir...), los más innovadores y cambiantes, en los que se mantiene una distancia interlingüística menor; es decir, donde más se aprecia la influencia de una lengua sobre la otra, que casi siempre es a través del préstamo reciente sobre todo.

Esto significa que, respecto de la renovación e innovación léxica, fuertes sobre todo en estos dos C.I. 2 y 10, el euskara aparece en fuerte dependencia lingüística en relación con el español.

4. Respecto del tipo de variables lingüísticas 4, de la corrección lingüística, o del grado de normatividad lingüística respecto del uso del léxico (ya que hemos considerado como correcta a la utilización lingüística normativa, siguiendo las directrices dadas sobre todo por Euskaltzaindia):

— Volvemos a encontrarnos reagrupados por una parte a los C.I. 1 (Partes del cuerpo humano) y 14 (Agricultura...), en que se da una mayor corrección lingüística o ajuste respecto de la norma lingüística, por otra parte esperable como veremos mejor más adelante.

— Por otra parte, los C.I. 2 (Prendas de vestir...) y 10 (Juegos...), donde encontramos mayor incorrección lingüística.

5. Respecto del tipo de variables lingüísticas 5, de las divisiones significativas o semánticas de interés propias de cada C.I. o mundo significativo amplio, y por tanto distintos ya en cada C.I.:

— En el C.I. 2 (Prendas de vestir...): la cantidad de producción léxica respecto de las prendas de vestir es claramente predominante respecto del relacionado con el calzado.

Esto puede deberse a un grado de diversificación o heterogeneidad significativa mayor en el mundo del vestido, es decir, a una mayor cantidad de conceptos distintos a expresar; pero puede deberse también a una mayor disponibilidad significativa, a una mayor importancia del vestido para nuestra población estudiada.

— En el C.I. 10 (Juegos...): si tenemos en cuenta que los deportes (SDP) son actividades a realizar sobre todo en la calle y no en casa, vemos que cuantitativamente hay un claro predominio del léxico producido relacionado con los juegos-diversiones-entretenimientos a realizar sobre todo en la calle

(SDP + 5CL), de hasta el 51,65%, frente al 28,25% obtenido para los juegos-diversiones-entretenimientos a realizar sobre todo en casa (5CA).

Es también en este C.I. 10 donde se da el mayor índice de 5RE o resto significativo, es decir, de producción léxica de difícil clasificación en el esquema significativo propuesto por nosotros por interés sobre todo sociolingüístico, y de cara a una planificación sociolingüística.

— Y en el C.I. 14 (Agricultura y plantas...), cuantitativamente: por una parte es claramente predominante el mundo de la ganadería y los animales; y por otra, los mundos de la agricultura y plantas, y de la pesca y peces, ocupan lugares parecidos entre sí, siendo el mundo de la pesca el menos representado.

## II

Desde el punto de vista del grado de homogeneidad o heterogeneidad con que se ha dado la producción léxica en relación con estas variables lingüísticas, en ambas lenguas, es decir, el grado de heterogeneidad con que nuestra población estudiada se ha comportado lingüísticamente en el uso de ambas lenguas, y según puede verse en el mismo cuadro dado más arriba:

1. Respecto del tipo de variables lingüísticas 1, de la cantidad de producción léxica: la dispersión es casi nula; es decir, los individuos de nuestra población estudiada se comportan casi del mismo modo, producen casi la misma cantidad de ocurrencias (expresiones o palabras).

2. Respecto de los tipos de variables lingüísticas 2: del grado de lexicalización y la clase léxica, y 5: de las divisiones significativas o semánticas propias de cada C.I., la dispersión es, en general, pequeña; es decir, nuestra población estudiada se comporta lingüísticamente de modo bastante homogéneo.

3. Respecto del tipo de variables lingüísticas 4, de la corrección lingüística: la dispersión es ya significativa. Esto quiere decir que frente a individuos cuya producción léxica es correcta, están los individuos cuya producción léxica es incorrecta.

4. Pero es sobre todo respecto del tipo de variables lingüísticas 3, de la distancia interlingüística entre el euskara y el español, donde se dan los mayores índices de dispersión, y ya índices de heterogeneidad grandes, y sobre todo respecto de los C.I. 2 (Prendas de vestir...) y 10 (Juegos...) otra vez. Es decir, es respecto de este grupo de variables lingüísticas donde nuestra población se comporta muy heterogéneamente.

La disminución de la distancia interlingüística entre ambas lenguas se hace fundamentalmente a costa del euskara, es decir, a través de parecerse

cada vez más el euskara al español, aumentando la dependencia léxica del euskara respecto del español, sobre todo a través de los préstamos recientes. Como respecto de estas cuestiones, además, es donde nuestra población se comporta muy homogéneamente en la producción léxica en español, esto quiere decir que se comporta muy heterogéneamente en la producción léxica en euskara, con índices bastante mayores en realidad para cada lengua que los que aparecen reflejados en el cuadro para ambas lenguas a la vez.

En resumen, respecto de este tipo de problemas ligados al uso léxico es, pues, donde se da el bilingüismo más desequilibrado en nuestra población, y por tanto donde más énfasis habría que hacer en una planificación lingüística del bilingüismo, de cara al bilingüismo social equilibrado.

También por otro camino, a través del método estadístico-informático de análisis multivariante, hemos llegado a las mismas conclusiones. Efectivamente, dicho análisis multivariante nos descubría que las variables lingüísticas más importantes para el estudio de los comportamientos lingüísticos distintos de nuestra población estudiada eran, también, las variables de los tipos: 3, de la distancia interlingüística, y 4, de la corrección lingüística, ya que siempre aparecen juntas definiendo el factor 1 (el más significativo), en los cuatro C.I., como veremos mejor más adelante.

### III

Desde el punto de vista de las correlaciones, o relaciones interdependientes, entre las variables lingüísticas de los distintos tipos de variables, nos hemos encontrado con lo que esperábamos, se nos ha confirmado una de nuestras hipótesis-punto de partida, ya que entre nosotros se da también lo mismo que en otras situaciones de lenguas en contacto, como por ej. en Canadá entre el francés y el inglés: de que a mayor distancia interlingüística mantenida, mayor grado de corrección en el uso lingüístico de ambas lenguas; y a la inversa, de que a menor distancia interlingüística mantenida (a través de los préstamos, de las interferencias, de la sustitución, etc.), mayores probabilidades de un uso incorrecto de ambas lenguas.

Así, en nuestro trabajo nos encontramos con que las variables lingüísticas pertenecientes a distintos tipos de variables, más fuertemente correlacionadas entre sí, y con un índice de correlación muy fuerte (porque ya desde el índice del 0,57 la correlación resulta significativa en nuestro trabajo), en los cuatro C.I., son:

— *3DI* (proporción de expresiones y de palabras fonológicamente diferentes en las dos lenguas), con *4BI* (proporción de expresiones y de palabras bien o correctamente utilizadas):

— En el C.I. 1 (Partes del cuerpo):  $r = 0,79$ ;

- En el C.I. 2 (Prendas de vestir...):  $r = 0,65$ ;
- En el C.I. 10 (Juegos...):  $r = 0,93$ ;
- En el C.I. 14 (Agricultura...):  $r = 0,87$ .

— *3PA* (proporción de expresiones y de palabras fonológicamente parecidas en ambas lenguas), con *4MA* (proporción de expresiones y de palabras mal o incorrectamente utilizadas):

- En el C.I. 1:  $r = 0,82$ ;
- En el C.I. 2:  $r = 0,64$ ;
- En el C.I. 10:  $r = 0,93$ ;
- En el C.I. 14:  $r = 0,86$ .

Es decir, nos encontramos con que estos dos tipos de variables lingüísticas: el 3 respecto de la distancia interlingüística, y el 4 respecto de la corrección lingüística, aparecen en nuestro trabajo como los más importantes e interesantes, tanto desde el punto de vista lingüístico como también desde el sociolingüístico (de cara a la caracterización lingüística de los distintos grupos o estratos de población distinguidos), y aparecen a la vez como los tipos de variables lingüísticas más fuertemente correlacionados entre sí.

A modo de ejemplo, añadimos un gráfico a continuación: el plano 1.<sup>o</sup> ofrecido por el ordenador para el análisis multivariante, respecto del C.I. 14 (Agricultura...), que representa gráficamente todo el conjunto de correlaciones de todas las variables lingüísticas estudiadas entre sí, del modo más significativo (Ver gráfico.)

Si nos fijamos en el factor 1 (en el eje horizontal del plano), el más significativo, vemos que viene definido sobre todo por el conjunto de variables *3DI* y *4BI* por un lado (muy fuertemente correlacionadas también entre sí, dada su gran proximidad), y por las variables *3PA*, *3SU* y *4MA* por el lado opuesto (también muy fuertemente correlacionadas entre sí).

Además, puede verse también el lugar que ocupan el resto de las variables lingüísticas en este espacio de correlaciones, y en relación con dicho factor 1. Así, por ejemplo:

— *5GN* (proporción de expresiones y de palabras que significan respecto de la ganadería y los animales, en el C.I. 14): aparece bastante fuertemente correlacionada con *3DI* y *4BI*;

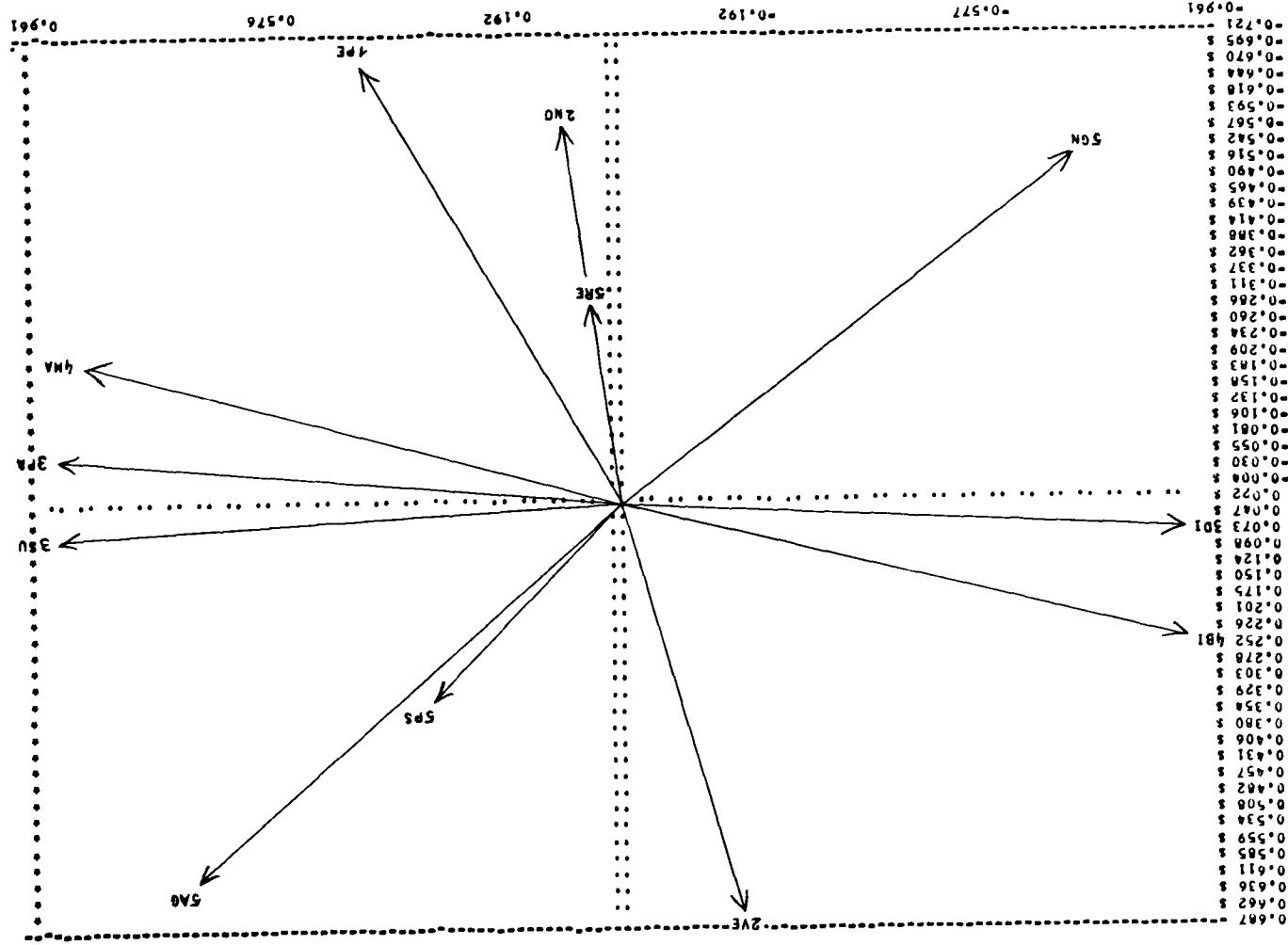
— *5AG* (proporción de expresiones y de palabras que significan respecto de la agricultura y las plantas) bastante correlacionada también, pero con *3PA* y *4MA*.

— Y *5PS* (proporción de expresiones y de palabras que significan respecto de la pesca y los peces), aunque más débilmente, correlacionada también con *3PA*, *3SU* y *4MA*.

Es decir, que el mundo de la ganadería y los animales ha sido expresado por nuestra población con un léxico que mantiene una distancia o diferencia-



ALGUNOS ASPECTOS DEL EUSKARA UTILIZADO HOY



ción interlingüística grande entre el euskara y el español, con pocos préstamos recientes y pocas sustituciones léxicas, y además con un léxico utilizado con gran corrección lingüística. Pero a la inversa, que el mundo de la agricultura y las plantas se ha expresado con un léxico que mantiene una distancia o diferenciación interlingüística escasa, y además que es utilizado muy incorrectamente; y que en el mundo de la pesca y los peces, aunque más débilmente, se refleja la misma tendencia que la vista respecto de la agricultura y las plantas.

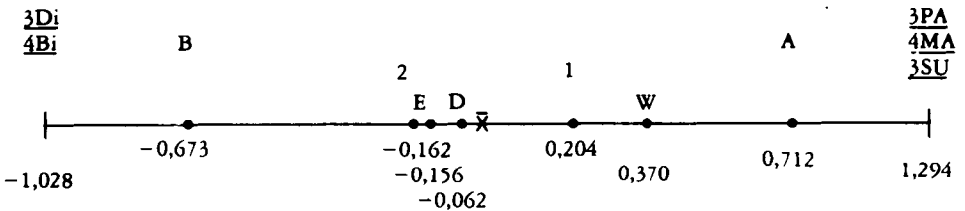
#### IV

Y desde el punto de vista del comportamiento lingüístico de nuestros bilingües estudiados (los escolares de 5.º y 8.º de EGB, de las Ikastolas de la Comarca de San Sebastián), en cada una de las dos lenguas en contacto: el euskara y el español, respecto de todo este conjunto de variables lingüísticas analizadas, nos encontramos también con lo que esperábamos: con una situación de bilingüismo social bastante o muy desequilibrada.

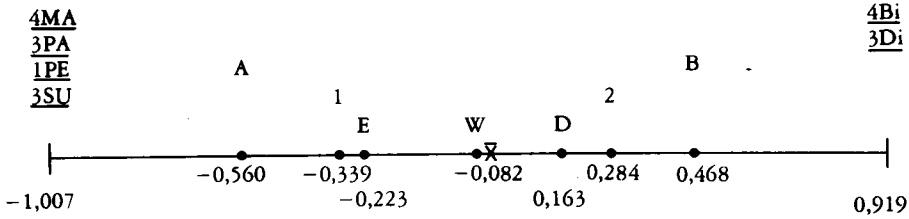
Teniendo en cuenta que los índices de valor utilizados hasta ahora reflejaban los valores de producción media de cada variable lingüística, respecto de la producción léxica en ambas lenguas a la vez, esto significaba que los valores reales de cada variable son bastante mayores cuando se refieren a la producción léxica en cada lengua por separado, como decíamos ya más arriba.

Para dar una idea rápida mayor del distinto comportamiento lingüístico de nuestros bilingües en relación al uso de cada una de sus dos lenguas en contacto, y para resumir el conjunto de conclusiones dado en este artículo, ofrecemos a continuación los cuatro gráficos siguientes, que resumen el comportamiento de los distintos grupos o estratos de población estudiados en relación a las variables lingüísticas más importantes, y por tanto las variables definidoras de los factores 1, en las salidas de ordenador para los cuatro C.I.:

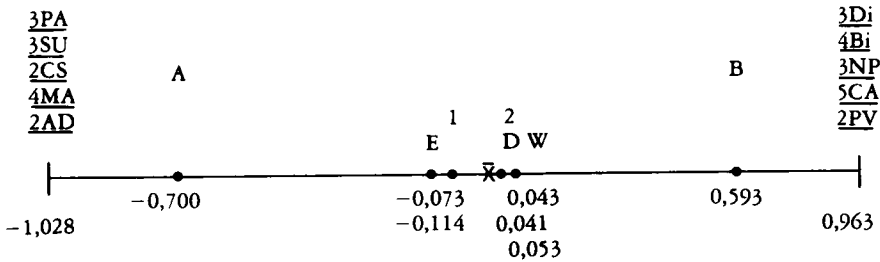
— C.I. 1 (Partes del cuerpo humano):



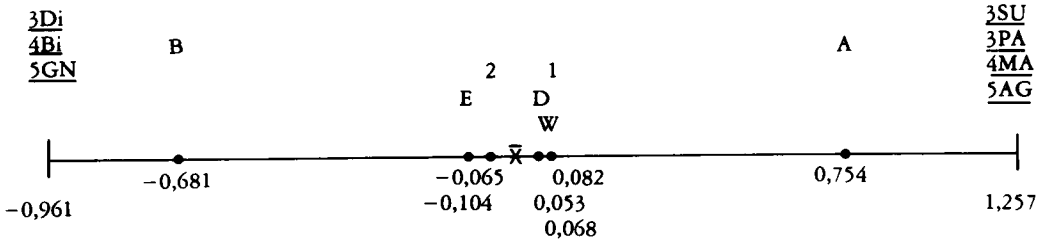
— C.I. 2 (Prendas de vestir y calzado):



— C.I. 10 (Juegos. Diversiones. Entretenimientos):



— C.I. 14 (Agricultura y plantas. Ganadería y animales. Pesca y peces):



1. Respecto de las variables lingüísticas:

— En los gráficos de los cuatro C.I. se pueden ver las variables lingüísticas que aparecen más fuertemente correlacionadas con el factor 1, y por tanto que aparecen definiendo dicho factor 1. Estas son las variables del tipo 3, de la distancia interlingüística, y del tipo 4, de la corrección lingüística. Es decir, se trata de las variables lingüísticas que mejor sirven para caracterizar a los distintos grupos sociolingüísticos; o a la inversa los distintos grupos sociolingüísticos se diferencian entre sí sobre todo por sus distintos comportamientos lingüísticos en relación a dichas variables lingüísticas.

— Además, en algunos C.I. aparecen también otras variables lingüísticas definiendo el factor 1, aunque siempre con menor fuerza que las variables de los tipos 3 y 4, es decir, con menores índices de correlación con dicho factor.

— Recordando que las variables que aparecen en cada uno de los polos o extremos definiendo un factor están significativamente correlacionadas entre sí también, y las primeras muy fuertemente correlacionadas, esto significa, por ejemplo:

Que en el C.I. 10 (Juegos...): a menor distancia interlingüística (+3PA, +3SU), mayor incorrección lingüística (+4MA); pero también, mayores probabilidades de que se den una mayor proporción de nombres utilizados en un caso determinado (+2CS), y una mayor proporción de adverbios (+2AD).

O a la inversa, a mayor producción léxica de 2CS y de 2AD, mayores probabilidades de +3PA, +3SU y +4MA. También, que a +3DI y +4BI, también +3NP (proporción de nombres propios), +5CA (proporción de juegos a realizar sobre todo en casa) y +2PV (proporción de expresiones o perífrasis formadas por preposición + verbo); o a la inversa.

Que en el C.I. 14 (Agricultura...): a +3DI y +4BI, también +5GN (proporción de expresiones o de palabras que significan respecto de la ganadería y de los animales); o a la inversa que a +5GN, mayor corrección lingüística y mayor distancia interlingüística. Y que a +3SU, +3PA y +4MA, +5AG (proporción de expresiones y de palabras que significan respecto de la agricultura y las plantas); o que a +5AG, menor distancia interlingüística y mayor incorrección lingüística.

2. Y respecto de los grupos sociolingüísticos distinguidos, formados a través de distintos reagrupamientos pero siempre de la misma población estudiada (el conjunto de los escolares de 5.º y 8.º de EGB de las Icastolas de la Comarca de San Sebastián), según estos distintos criterios:

Por la geografía:

*D* (Donostia/San Sebastián), que representa a todo el léxico producido en ambas lenguas, y por los dos grupos de edad, en la ciudad de Donostia;

*E* (zona Este de la Comarca: Errenteria, Pasaia, Oiartzun), que representa a todo el léxico producido en dicha zona Este en ambas lenguas, y por los dos grupos de edad;

*W* (zona Sur-Oeste: Hernani, Urnieta, Lasarte, Usurbil), que representa a todo el léxico producido en dicha zona Sur-Oeste.

Por la edad:

1 (5.º de EGB), que representa a toda la producción léxica de los alumnos de 5.º EGB en ambas lenguas y en toda la comarca;

2 (8.º de EGB), que representa a todo el léxico producido por los alumnos de 8.º de EGB en ambas lenguas, y en toda la Comarca.

Por la lengua utilizada:

*A* (el euskara), que representa a todo el léxico producido en euskara en toda la Comarca, y por los alumnos de ambos grupos de edad;

*B* (el español), que representa a todo el léxico producido en español en toda la comarca y por ambos grupos de edad:

— El comportamiento lingüístico más constante en los cuatro C.I., y también el más diferenciado entre sí, se da entre los grupos sociolingüísticos *A* (producción léxica total en euskara) y *B* (producción léxica total en español).

La utilización de cada una de las dos lenguas en contacto, por parte de nuestros alumnos bilingües, es lo que mejor discrimina o clasifica a nuestra población estudiada. Es decir, los comportamientos lingüísticos más distintos se han debido sobre todo a la lengua utilizada, y no tanto a las diferencias de edad o de zona geográfica. Por eso, *A* y *B* son los grupos sociolingüísticos mejor representados en el factor 1 (los más a los extremos de dicho factor).

*A* y *B* aparecen colocados con índices altos en el factor. Pero *A* (producción léxica en euskara) siempre ligada a las variables de la menor distancia interlingüística y de la mayor incorrección lingüística; mientras que *B* (producción léxica total en español), por el contrario, y también con mucha diferencia, siempre viene ligada a las variables de la mayor distancia interlingüística y de la mayor corrección lingüística. Esto significa que, respecto de estas variables lingüísticas, nos encontramos con una situación de bilingüismo social muy desequilibrado, en favor del español y en perjuicio del euskara.

— El resto de los grupos sociolingüísticos, tanto en razón a las diferencias de edad como a las distintas zonas geográficas, están débilmente caracterizados en relación a estas variables lingüísticas de estos factores 1.

A pesar de ello, puede apreciarse el comportamiento lingüístico también constante de los grupos de edad, ya que 1 (producción léxica total de los alumnos de 5.º EGB) siempre aparece en los factores 1 relacionado con las variables lingüísticas de la menor distancia interlingüística y de la mayor incorrección lingüística; y por el contrario, 2 (producción léxica total de los alumnos de 8.º EGB) siempre relacionado con la mayor distancia interlingüística y la mayor corrección lingüística. Es decir, con el aumento de la edad (de 5.º a 8.º de EGB) se da un aprendizaje lingüístico, fundamentalmente respecto del euskara, una recuperación cualitativa del euskara, aunque con distintos grados de intensidad en cada C.I. (en nuestro trabajo, sobre todo respecto del C.I. 2: prendas de vestir y calzado).

Y respecto de los grupos geográficos, la situación es ya más compleja, y más difícil de sintetizar.