

ESTUDIO DE LA COMPETENCIA SOCIOLINGÜÍSTICA: LOS MODELOS PROBABILÍSTICOS *

0. Ninguna escuela lingüística ni ningún pensador independiente han negado jamás la existencia de la variación lingüística. Las gruesas diferencias que se observan en el corpus teórico del siglo xx producido por esta ciencia con respecto al papel desempeñado —si alguno— por la variación en el diseño de modelos, obedece sin duda al grado de importancia que los teóricos le confieran en la conformación de las gramáticas. La dicotomía que muy tajantemente se observa está basada, por un lado, en la idea de que la variación es un subproducto externo, ajeno al sistema mismo, producto de una actualización asistemática y caprichosa —*parole, performance*— de un sistema regulado e invariable —*langue, competence*— al que miran como objeto primario, si no único, de estudio, y por otro lado, en la concepción de que la variación debe ser incorporada sistemáticamente a la descripción y a la teoría lingüística, puesto que es parte integrante del sistema.

El primer punto de vista, apoyado en el supuesto carácter errático, indiscriminado, y en todo caso superficial, de la variación no ha recibido corroboración empírica en los más recientes estudios sociolingüísticos. Por el contrario, la hipótesis desplegada con éxito por esta disciplina dice que la variación que se observa en la actuación lingüística de una comunidad determinada es el resultado de unos procesos

* Comunicación plenaria presentada al décimo Simposio de la Sociedad Española de Lingüística en diciembre de 1980. El presente trabajo respeta el esquema y el texto de la comunicación, pero incorpora una bibliografía esencial, tanto de obras expresamente citadas como de otras cuya importancia para el tema tratado es sobresaliente.

armónicos y condicionados por factores tanto lingüísticos como extralingüísticos. Supone esta hipótesis que la variación completamente libre o el polimorfismo, injustificado en apariencia, son circunstancias que al recibir análisis detenido dejan al descubierto el cumplimiento sistemático de ciertas reglas, de donde se desprende que la variación es parte integrante de la estructura lingüística de todo diasisistema. Los factores que impulsan o determinan esta variación son, además de los propiamente lingüísticos, de carácter social.

Pretender que la variación observada en la actuación lingüística del hablante, y consiguientemente del grupo, sea totalmente fortuita nos llevaría a postular —como se ha hecho en más de una ocasión— que nada regularía, por ejemplo, el que un sujeto de Las Antillas realizara el segmento fonológico *-r/* como un sonido vibrante, como uno lateralizado, como uno fricativizado o inclusive como un cero fonético, es decir, que estaríamos ante una «variación libre» en el sentido bloomfieldiano de la expresión, o ante el cumplimiento indiscriminado de reglas opcionales, según el modelo generativo. Sin embargo, desde la investigación pionera de William Labov sobre la contracción y la elisión de la cópula en inglés se viene comprobando el hecho contrario: la variación lingüística es sistemática y está altamente correlacionada con factores lingüísticos y sociales. A partir de aquí se ha comprobado reiteradamente que la actuación no está libre de covariación y que existe una dependencia sistemática entre frecuencia de ejecución de determinada regla y determinados factores condicionantes.

Explicar la lateralización, por una parte, por otra la fricativización y elisión del segmento *-r/* como un simple proceso de selección entre varias alternativas posibles es una dramática simplificación de la realidad. En los sociolectos de San Juan, por ejemplo, son varios los condicionantes, tanto lingüísticos como extralingüísticos, que impulsan o retienen el cumplimiento de estas reglas: distribucionales, contextuales y funcionales. El Cuadro 1 indica claramente que en posición interna el contexto pre-obstruyente favorece la fricación y la lateralización. La posición pre-lateral impulsa sobre todo las lateralizaciones, pero se trata en realidad de casos de asimilación, esperables por lo demás, dada la insistencia con que otras investigaciones apuntan este hecho. Las elisiones son aquí de cierta consideración, sobre todo si se repara en lo escaso del fenómeno en los dialectos puertorriqueños.

	obstruyente	lateral	nasal	vocal	pausa
<i>/r/ interna</i>					
[l]	34.6	40	17.4		
[r]	14.8	27.2	19.3		
[r̄]	43.4	22.7	61.8		
[ø]	7	9	1.2		
N	3,596	22	475		
<i>/r/ final monomorfémica</i>					
[l]	56.5	37.7	39.1	13.5	55.1
[r]	12.3	11.6	2.1	19	12
[r̄]	28.8	45	58.6	65.7	29.7
[ø]	2.1	5.5	0	1.3	3.1
N	599	180	46	383	225
<i>/r/ final infinitiva</i>					
[l]	37.1	37	38.6	25.4	37.6
[r]	11.1	12.7	5.1	16.8	10.3
[r̄]	47.9	42.3	49	50	45.8
[ø]	3.8	7.8	7.1	7.1	6.1
N	1,131	586	251	1,019	569

CUADRO 1: Distribución de las variantes de */r/* según la distribución, el contexto fónico y el *status* gramatical

En posición final de palabra, tanto en los casos no gramaticales como en los gramaticales, las obstruyentes siguen favoreciendo esencialmente la lateralización y la fricación, aunque con diferente acento. En los contextos pre-laterales no se dan las asimilaciones con la fuerza de la posición interior; en ambos casos, la variable de mayor frecuencia es la fricativa. Con respecto al efecto de la nasal pospuesta, se sigue el patrón ya visto para la */r/* interna: preponderancia de la fricación, aunque las lateralizaciones no sean desestimables. El contexto pre-vocálico es un claro impulsor de la variante fricativa, mientras que el pre-pausal tiende a favorecer a la variante [l], si se trata de */r/* monomorfémica.

Cuando se examinan, además, los factores sociales se observa la existencia de algunas relaciones asociativas fuertes, de otras menos fuertes y de claras disociaciones.

		[l]	[r]	[r']	[ø]
Sexo	M	17.8	6.4	20	2.4
	F	16.6	7.5	25.6	3.1
Edad	I	10.7	4.7	16.8	1.4
	II	17	6.2	19.6	3
	III	6.8	3	9.1	1.1
Nivel Socio-económico	1	2.6	9.6	6	.3
	2	9.3	1.6	12.1	.9
	3	7.6	1.4	11.7	1.4
	4	15	1.4	15.8	3.1
Procedencia	A	20.4	11.7	30.3	3.4
	B	2.5	.5	3.7	.5
	C	4.3	1	2.9	.2
	CH	7.2	.7	8.6	1.4
N = 9,082					

CUADRO 2: Correlación entre variantes de /r/ y factores sociales.

Tanto los hombres como las mujeres realizan más variantes fricativas que ninguna otra, pero los hombres lateralizan más; las mujeres en cambio tienen más vibrantes y también más elisiones. Esto último parece ser la excepción al comportamiento lingüístico más conservador del sexo femenino, mientras realizan más variantes estándar y menos lateralizaciones, fenómeno este último estigmatizado en San Juan, muestran más frecuencia, aunque no mucho mayor (.7) de ceros fonéticos.

Las generaciones siguen el mismo esquema en cuanto a preferir abiertamente la variante fricativa, pero el resto de los datos no indica nada especial. El nivel socio-económico en cambio, deja ver cómo la lateralización, la fricativización y la elisión aumentan a costa de la vibrante, a medida que descendemos en el espectro. En cuanto a la procedencia no hay datos dramáticos que subrayar; los sujetos de procedencia urbana metropolitana tienen la fricativa como variante

preferida, seguida de las laterales. El esquema se mantiene menos en la procedencia rural C donde la lateralización es la realización dominante. Obsérvese que en CH, también procedencia rural, aunque la fricativa encabeza las variantes, la lateral se encuentra a muy poca distancia de aquella.

Estos datos, entre otros muchos que podrían aducirse, sirven de ejemplo a lo apuntado con anterioridad: la variación es sistemática y está condicionada por factores lingüísticos y extralingüísticos.

0.1. Dos interrogantes importantes se abren entonces al investigador: 1) ¿Cómo estudiar la variación?, y 2) lo que es más importante, si aceptamos la premisa de que ésta es parte integrante de la estructura lingüística y no un mero accidente casual, ¿de qué forma podría ser incorporada a un modelo explicativo riguroso y aceptable?

Los intentos por estudiar científicamente la variación lingüística son varios y su explicación detallada sería objeto de una presentación particular; baste mencionar aquí los nombres de Bickerton, de Camp, Elliot, Legun y Thompson, Ross, Sag, Bailey, además del grupo que estudia la variación utilizando instrumentos probabilísticos y que incluye a Labov y a toda una serie de estudiosos comandados por Sankoff, Cedergren y Rousseau, todos del círculo lingüístico-matemático de Montreal. Lo más importante de estos últimos es que además de ofrecer sólidos fundamentos estadísticos para el análisis de la variación, circunstancia ésta que los separa de los demás variacionistas, intentan crear un modelo gramatical coherente que la incorpore.

Las bases para tal modelo son simples, ya que procuran insertarse en una concepción muy trabajada, ampliamente difundida y aceptada por todas partes: la generativo-transformacional. Partiendo de aquí puede repostularse que si toda actuación lingüística no es más que el reflejo, por aproximado e imperfecto que sea, de la competencia, parece obligado concluir que los aspectos de la actuación que funcionan sistemáticamente son también reflejo de la competencia lingüística. Se lanza la hipótesis de un nuevo concepto de competencia, más abarcador, que añada a la competencia gramatical ortodoxa otros aspectos inclusive los sociales.

Este último postulado ha dividido en dos el pensamiento lingüístico actual con respecto al papel de la variación en los modelos teóricos. Para algunos sociolingüistas la variación es también parte inte-

gral de la competencia y no mero problema de inventario externo relacionado con lo social, lo dialectal, lo situacional y lo temporal; no es, por consiguiente, explicable acudiendo a varias gramáticas distribuidas a lo largo de un continuum, u organizada mediante algún otro artificio, como sugieren inclusive ciertos variacionistas. La actuación sería un reflejo estadístico de la competencia y su formalización tendría por base la llamada «regla variable».

0.2. Una gramática dada puede permitir que R sea reescrito como BA, pero sólo, por ejemplo en los casos en que esa R se encuentre inmediatamente después de T y antes de U

$$(1) R \longrightarrow BA/ T \text{ ______ } U$$

y también que R sea reescrita como U cuando le sigue un contexto XY

$$(2) R \longrightarrow U/ \text{ ______ } XY$$

Esto obliga a considerar dos factores importantes: contexto condicionador y orden de aplicación de las reglas, pues el decurso TTRRXYZ puede ser reescrito como TTRUXYZ según la regla (2), y después como TTBAUXYZ según la (1).

Este ejemplo deja en claro que toda gramática condicionada por el contexto tiene que incluir reglas determinísticas o probabilísticas que le permitan seleccionar en cada etapa de la derivación cuál de los símbolos va a ser considerado para su reescritura. Una diferente selección de reglas desemboca generalmente en resultados diferentes, sobre todo si se considera que la aplicación de algunas producen un nuevo contexto cuya descripción estructural impulsaría el cumplimiento de otra.

Un punto más debe ser tenido en cuenta y es que las reglas (condicionadas por el contexto o no) suelen presentarse como un haz de reglas similares, que los recursos de la formalización generativa condensa siempre. Así en lugar de

$$(3) \begin{array}{l} R \longrightarrow BA/ T \text{ ______ } KG \\ R \longrightarrow BA/ U \text{ ______ } KG \\ R \longrightarrow BA/ T \text{ ______ } LH \end{array}$$

y otras 33 reglas más, por ejemplo, que podrían ser añadidas a esta nómina, la gramática ofrece la siguiente:

$$(4) R \longrightarrow BA / \left\{ \begin{array}{c} T \\ U \\ V \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} K \\ L \\ M \\ N \end{array} \right\} \left\{ \begin{array}{c} G \\ H \\ I \end{array} \right\}$$

Lo importante aquí es subrayar que estimar por separado las 36 posibilidades de (4) es una tarea inadecuada, tanto lingüística como estadísticamente, porque la presencia de un símbolo específico en una determinada posición en el decurso puede hacer variar sistemáticamente la posibilidad de aplicación de la regla. Lo que se impone es un análisis estadístico del efecto de los varios símbolos presentes en el decurso en el que se aplica una regla opcional. Hay que advertir a renglón seguido que las llamadas reglas opcionales, sea cual sea el marco teórico en el que se inserten, se convierten siempre, desde el punto de vista probabilístico, en reglas condicionadas por el contexto, pues aunque un símbolo sea susceptible de ser reescrito independientemente del contexto, la regla se cumplirá más frecuentemente en unos contextos que en otros. La regla de lateralización en los sociolectos de San Juan se cumple con frecuencia mayor en posición final de palabra que en posición interna, y allí los contextos influyen en el orden jerárquico obstruyente > pausa > nasal.

La regla variable conlleva la indicación probabilística de su aplicación, según estén o no presentes en el decurso factores condicionantes específicos, ofreciendo cada uno de ellos su particular contribución al cumplimiento o bloqueo de la misma. De aquí se deduce que en los sociolectos de nuestro ejemplo *-r/* tiene más probabilidad de lateralizarse en el decurso ... *el parador que visitamos*, que en ... *el mar ahora está alborotao*, aunque la regla de lateralización sea aparentemente de carácter opcional no condicionada por el contexto fónico.

Para calcular estas probabilidades se parte de la frecuencia observada en un corpus representativo de la actuación lingüística de los hablantes de una comunidad dada. Un procedimiento matemático convierte esa frecuencia real en probabilidad teórica, y este resultado pasa a convertirse en un índice probabilístico de aplicación; cada factor o familia de factores presentes en el decurso tendrá el suyo. Desde 1969 hasta hoy se han diseñado diferentes modelos probabilís-

ticos en un empeño por refinar la metodología de las reglas variables y consecuentemente por acercarse más a una descripción y explicación adecuadas de la competencia sociolingüística.

1. El primer modelo fue preparado por Labov en 1969 y usado con éxito por el mismo investigador en su estudio de la contracción y elisión de cópula en el inglés de la ciudad de Nueva York.

El análisis de su corpus le indicaba que la contracción de la cópula era casi obligatoria para todos los hablantes si la FN sujeto era un pronombre (*He is going* → *He's going*) lo que ofrecía por lo tanto poco interés para el estudio de la variación. Si la FN sujeto no era pronombre, entonces la mayor frecuencia de contracciones se daba cuando la FN sujeto terminaba en vocal

Martha is going → *Martha's going*

El examen del elemento contextual pospuesto a la cópula revelaba un orden decreciente de contracciones según la siguiente jerarquía:

- a) ante el futuro *going to* (> *gonna*): *John's gonna go*
- b) ante otro tipo de gerundio: *John's eating*
- c) ante un adjetivo predicativo o un locativo: *John's happy*
John's in the bathroom
- ch) ante una FN: *John's a good man*

Las generalizaciones que pueden efectuarse de estos datos de actuación dicen que la contracción resulta afectada por el contexto fónico precedente y la clase de segmento que le siga. Estos hechos pasarán a traducirse, por ejemplo, en la siguiente regla:

$$V \longrightarrow \emptyset / \left\langle \begin{array}{c} +V \\ -V \end{array} \right\rangle_a \# \# \left[\text{---} \right] C \# \# \left\langle \begin{array}{c} +Vb \\ -Vb \end{array} \right\rangle_b$$

aunque por el momento desconozcamos las probabilidades específicas de los elementos variables, que aquí van entre corchetes angulares.

El Cuadro 3 pone de manifiesto los datos básicos para iniciar el análisis probabilístico¹.

¹ Las cifras que utilizo en este ejemplo no son reales sino ficticias de manera de facilitar la inteligencia del texto; las tomo directamente del artículo de P. Kay y Ch. K. McDaniel, pág. 9. Este préstamo, motivado sólo por circunstancias didácticas, no me hace identificarme con el objetivo y las argumentaciones de los autores, con quienes discrepo en lo esencial.

verbo pospuesto	vocal antepuesta	
	sí	no
sí	99/110	91/130
no	109/218	36/120

A) Frecuencias observadas

verbo pospuesto	vocal antepuesta	
	sí	no
sí	.9	.7
no	.5	.3

B) Proporciones (ratio)

CUADRO 3: Frecuencia y proporciones de la regla de contracción de cópula según el factor fónico antepuesto y el segmento pospuesto.

Los diferentes entornos muestran un influjo diferente sobre la posibilidad de que se cumpla la regla de contracción. Una vocal precedente patrocina la contracción: $.9 > .7$; $.5 > .3$, y la FV pospuesta también: $.9 > .5$; $.7 > .3$.

Pero este tipo de cuantificación es superficial en extremo; lo que se necesita es producir un conjunto de contribuciones numéricas correspondientes a cada rasgo o factor condicionante y un mecanismo algebraico que combine los pesos correspondientes a cada celda particular, de modo que sea posible predecir su frecuencia de realizaciones y que éstas se correspondan con las frecuencias observadas. Si las frecuencias estimadas reflejan las frecuencias reales con una fidelidad razonable, el procedimiento empleado tendrá una buena razón, aunque no la única, para mantenerse.

La ecuación utilizada por Labov para combinar los pesos de los diferentes elementos condicionantes variables, sólo conlleva dos tipos de operaciones, adición y sustracción, por lo que el modelo ha sido llamado aditivo.

El modelo aditivo traduce la premisa de que la probabilidad de aplicación de una regla en un contexto dado es simplemente la suma de un número de cantidades, una por cada rasgo relevante en el contexto:

$$p = p_0 + \alpha_i + \alpha_j + \dots$$

La probabilidad resultaría de adicionar los valores de cada uno de los rasgos i, j , etc. de estar éstos presentes en el entorno lingüístico, y de un input (p_0), común a todos los contextos. El efecto de un determinado rasgo depende así solamente de su presencia en el contexto y no de otros aspectos del mismo.

El modelo está basado en el análisis de variancia (ANOVA) ya que trabaja con cada una de las posibilidades consideradas en el diseño y mide la frecuencia relativa de la aplicación de la regla en cuestión bajo todas las posibles combinaciones de símbolos. El análisis de variancia da entonces estimados del efecto de cada condicionante en la probabilidad de que la regla se cumpla. Este tipo de análisis tro-pieza, sin embargo, con una dificultad de envergadura al trabajar con datos lingüísticos; las gramáticas rechazan muchas de las posibilidades combinatoriales teóricas que simplemente no generan por anti-gramaticales; esto causa que muchas celdas del cuadro de datos aparezcan vacías y que por lo tanto sean distribuidas de forma muy desequilibrada, y que aun en las celdas llenas los datos lingüísticos tengan muy diferente número de situaciones, tanto, que varían entre 1 y cientos.

Aunque el modelo laboviano supera estos inconvenientes, quedan todavía en pie algunas deficiencias. Cedergren y Sankoff han especificado tres criterios por los que debe medirse la bondad de todo modelo probabilístico. Primero, debe ajustarse bien a los datos sobre los que ha sido elaborado; segundo, debe ser aplicado a una amplia clase de reglas, y tercero, debe ser lingüísticamente significativo.

El modelo aditivo cumple con el primer criterio; sin embargo, falla en los otros. Es, sin duda alguna, aplicable a muchas clases de reglas, pero no a otras, en las que su aplicación predice probabilidades que sobrepasan los límites significativos de 0-1. Sería necesario proceder con ciertas modificaciones al modelo para que cuando las frecuencias observadas fueran muy diferentes en diversos contextos o cuando existiera un gran número de contextos diferentes no se produjeran tales resultados. El «ordenamiento geométrico fuerte», sugerido por el mismo Labov iba encaminado a impedir la existencia de cifras fuera del intervalo adecuado, actuando sobre el tamaño de los efectos de los varios rasgos. Otra forma de modificar el modelo aditivo es mediante el procedimiento de truncar o podar los resultados probabilísticos que excedan de 0-1, dejando en 1 las cifras excesivas

y en 0 las inferiores a esta cifra. En cuanto al tercer punto, este modelo no conduce a una interpretación simple de la naturaleza del componente probabilístico de la competencia lingüística.

2. El modelo multiplicativo, creado por Cedergren y Sankoff y aplicado por la primera en su investigación sociolingüística de la ciudad de Panamá, conserva algunas propiedades positivas del modelo aditivo, pero en general las enmiendas lo hacen superior de acuerdo a cada uno de los tres criterios señalados.

Cada factor está asociado con un efecto fijo, pero en lugar de sumarlos, se multiplican para llegar a conocer el valor pertinente a cada entorno particular. La formulación matemática sería:

$$P_{ij \dots k} = P_0 \times P_i \times P_j \dots \times P_k$$

que no es más que la interpretación probabilística de

$$X_{ij \dots k} = \log P_{ij \dots k}$$

aunque también es posible trabajar, y así parecen requerirlo algunas reglas, con el modelo negativo, el multiplicativo probabilístico de no aplicabilidad, cuya fórmula estaría compuesta por una serie de factores negativos, dados por la resta de la probabilidad de contribución de cada factor de 1:

$$(1 - p) = (1 - p_0) \times (1 - p_i) \times (1 - p_j) \times \dots (1 - p_k)$$

donde p_0 sigue siendo el input común a todos los entornos, e i, j, \dots, k los factores que contribuyen a la regla en cuestión, y cuyo valor oscilará siempre entre 0 y 1.

Supera este modelo al aditivo, primero, en que refleja muy puntualmente los datos recogidos en el análisis de la actuación sociolingüística; sin embargo, debe señalarse que utilizando sólo este criterio hay muchas ocasiones en que apenas habría motivo para preferir cualquiera de las dos versiones, positiva y negativa, del modelo multiplicativo sobre el aditivo. Con respecto al segundo criterio, la universalidad del modelo, es decir, al hecho de que sea siempre directamente aplicable, la superioridad de este nuevo modelo es indiscutible, puesto que no requiere de mecanismos adicionales para predecir sólo probabilidades *bona fide* entre 0 y 1. Sin duda lo más importante del mo-

delo es que si interpreta adecuadamente la naturaleza probabilística de la competencia; en este sentido conjuga lo descubierto por Labov sobre la independencia de la contribución de cada rasgo y la noción de independencia en el sentido probabilístico. Además, el combinar condicionantes variables se convierte en una extensión natural de los conceptos de ordenamiento conjuntivo o disyuntivo, según las probabilidades sean de aplicación o de no aplicación.

Para explicar el problema de la independencia y de la interacción los autores del modelo ofrecen un ejemplo clarificador. Como en la teoría de la probabilidad, para que un número de elementos sea mutuamente independiente se requiere que la probabilidad de todos ellos sea exactamente igual al producto de las probabilidades de todas sus ocurrencias individuales, si en un determinado día la probabilidad de lluvia es de $1/3$ y la de vientos intensos es de $1/2$, entonces los dos accidentes meteorológicos resultan independientes entre sí, sólo si las posibilidades de lluvia y de vientos se multiplican

$$1/3 \times 1/2 = 1/6$$

Para preparar el modelo multiplicativo de aplicación se parte de la base de que cada rasgo (i) permite la aplicación de la regla con una probabilidad p_i , en lugar de pensar que cada uno de ellos es capaz de hacer cumplir la regla. Son dos los elementos que llevan a la fórmula $p = p_i \times p_j \times \dots$, 1) la noción de independencia y 2) la condición de que todos los rasgos del entorno permitan la aplicación de la regla. En dicha fórmula hay una probabilidad p_k por cada rasgo k presente en el entorno, más el input probabilístico p_0 . Se trata de una generalización del ordenamiento conjuntivo de condicionantes, donde cada rasgo i en una condición ordenada conjuntivamente se considerará con valor $p_i = 1$, y cada rasgo k , incompatible con la condición, será $p_k = 0$.

Nótese que en los modelos multiplicativos, la contribución del rasgo i ($= p_i$) a la probabilidad de aplicación de la regla es en sí interpretable como una probabilidad, cosa que contrasta con el modelo aditivo donde no era posible interpretar cada coeficiente como probabilidad.

En su análisis de los sociolectos panameños, Cedergren encontró que los 22,167 casos de $-/s/$ se distribuían entre tres realizaciones de

superficie; una de ellas, la aspiración, con un 41 % del total de ocurrencias. Tras examinar los contextos en que ésta se producía, elaboró y postuló la siguiente regla variable:

$$s \rightarrow \langle h \rangle / \left\langle \begin{array}{c} \text{Nom} \\ \text{Adj} \\ \text{Det} \end{array} \right\rangle \left\langle \begin{array}{c} \text{plural} \\ \text{2a pers} \end{array} \right\rangle \langle \# \# \rangle \left\langle \begin{array}{c} + \text{ cons} \\ + \text{ voc} \\ - \text{ seg} \end{array} \right\rangle$$

La regla indica que los factores condicionantes de la aspiración, responsables por la frecuencia de su cumplimiento son: la naturaleza del segmento fonológico siguiente, la existencia de límite de palabra, el status gramatical de *s*, el tipo de sufijo y la forma gramatical co-ocurrente con el plural nominal. El Cuadro 4 muestra las frecuencias reales de la aspiración contra el total de casos de *-s/* y, entre paréntesis, los valores calculados

	<u>C</u>	<u>V</u>	<u>P</u>
Interna	$\frac{34}{1,791}$ (30)		
Monomorfémica	$\frac{89}{1,701}$ (67)	$\frac{157}{976}$ (174)	$\frac{153}{500}$ (174)
Verbo	$\frac{3}{30}$ (2)	$\frac{5}{13}$ (5)	
Determinante	$\frac{22}{862}$ (61)	$\frac{89}{179}$ (57)	
Adjetivo	$\frac{3}{142}$ (5)	$\frac{13}{59}$ (8)	$\frac{13}{68}$ (19)
Nombre	$\frac{36}{754}$ (22)	$\frac{69}{498}$ (67)	$\frac{146}{544}$ (142)

CUADRO 4: Frecuencias observadas de *h* y valores calculados, entre paréntesis.

La autora subraya el hecho evidente de que los datos ofrecidos en este cuadro señalan la sistematicidad con que el modelo captura los hechos de actuación. Y así es, sobre todo si se tiene en cuenta que sólo nueve parámetros han sido estimados para dedecir 28 celdas en el cuadro. Sin embargo, hay una excepción; en el caso de los determinadores las frecuencias esperables no se corresponden con las ob-

servadas: el modelo ha sido muy generoso en el contexto preconsonántico y muy conservador en el prevocálico. La conclusión obligada es el reconocimiento de que los determinadores no actúan independientemente, lo que constituye un reto para el modelo. En efecto, en algunos casos no hay independencia de interacción entre los factores contextuales de una regla. Aquí no se trata de un factor fonológico único —contexto prevocálico— sino de dos: al elemento vocálico se suma el acento ($\acute{v} = 70\%$, $v = 17\%$). No cabe duda de que este rasgo debe ser incorporado al contexto estructural de la regla de aspiración.

En lo que respecta al modelo, debe señalarse que dispone de posibilidades para manejar estas situaciones, pues estos casos de interacción pueden ser medidos añadiendo un grupo adicional de factores que represente la intersección de las dos categorías concernidas. Es más, este modelo sirve de ayuda para descubrir aquellos condicionantes lingüísticos que actúen interdependientemente, como en el presente caso panameño.

Otra conclusión importante que se desprende del análisis de los datos es que los condicionantes lingüísticos no interactúan con los sociales, pues los coeficientes de probabilidad asignados a los rasgos lingüísticos no ofrecen cambios significativos en los diferentes análisis. Esta independencia es una característica importante de la relación entre estructura lingüística y estructura social, ya que destruye la hipótesis de que en el eje sociolingüístico se cambia el peso de los diferentes rasgos lingüísticos que intervienen en una regla; la configuración de los condicionantes permanece relativamente fija; es el *input* probabilístico el que cambia de grupo en grupo sociolingüístico e inclusive de sujeto en sujeto.

3. Los modelos multiplicativos, positivo y negativo, fueron sustituidos pronto por los modelos logísticos. El primero de ellos fue elaborado por Sankoff en 1975, y desde entonces ampliamente utilizado por varios investigadores; es el que sirve de base a mi estudio sociolingüístico de San Juan.

La forma básica es ahora

$$\frac{p}{1-p} = \frac{P_0}{1-P_0} \times \frac{P_i}{1-P_i} \times \frac{P_j}{1-P_j} \times \dots$$

donde la probabilidad de las reglas está dada por la multiplicación de una serie de factores complejos, cuyos miembros son identificables con los integrantes de las fórmulas anteriores. El nuevo programa conserva todas las virtudes de los anteriores multiplicativos cuando p y p_i están cerca de 1 y cuando lo están de 0, y tiene también la simetría y la naturaleza aditiva del modelo de Labov cuando los parámetros están cerca de 1/2. El parámetro p_i sigue aquí interpretándose como la probabilidad de que la regla se cumpla cuando ningún otro rasgo (o *input*) la afecte; sin embargo el principio que combina las probabilidades individuales independientes ya no es más el simple de independencia estadística. La lógica del programa no permite que los parámetros se salgan del intervalo 0-1. Si el valor del parámetro p_i es 1/2, éste no tendrá efecto alguno sobre la regla; los que sobrepasen el valor de 1/2 favorecerán su cumplimiento y los inferiores a 1/2 la bloquearán.

Teniendo presentes estos puntos, revítese el cuadro probabilístico que acompaña a la regla de lateralización de *-r/* en los sociolectos de San Juan:

[int]	.46								
[— gram]	.51	[+ gram]	.51						
[obs]	.56	[lat]	.56	[nas]	.43	[voc]	.34	[-seg]	.60

CUADRO 5: Probabilidades de aplicación de la regla de lateralización en San Juan.

La posición interna no favorece la regla (menos de 1/2) la posición final sí, pero tanto el estatus gramatical como el monomorfémico tienen el mismo peso; de los contextos fónicos, el pre-pausal es el más favorecedor (.60), seguido del preobstruyente. Si se examina ahora el Cuadro 6, se verá el impacto de los factores sociales en el cumplimiento de la regla: el sexo masculino favorece su cumplimiento, en situación nada sorprendente estando la lateralización como está fuertemente estigmatizada en Puerto Rico y sabiendo, como sabemos, que las mujeres son más conservadoras en materia lingüística y más conscientes de la valoración social de los fenómenos del lenguaje; la generación más joven no favorece la aplicación de la regla, ni tampoco el nivel socioeconómico alto, ni la procedencia rural.

SEXO	EDAD	N. SOCIO- ECONÓMICO	PROCEDENCIA
M .53	I (20 — 34) .46	1 (alto) .26	A (urbana) .62
F .46	II (35 — 54) .52	2 .57	B .41
	III (55 +) .51	3 .53	C .44
		4 (bajo) .63	CH (rural) .50

CUADRO 6: Probabilidades de aplicación de la regla de lateralización de -/r/ en San Juan según factores sociales.

4. El segundo modelo logístico, preparado por Rousseau y Sankoff en 1978 y utilizado por Poplack en 1979 en su estudio sobre el dialecto puertorriqueño de Filadelfia, está basado, naturalmente, en el anterior, pero la gran ventaja que ofrece sobre aquél es que incorpora explícitamente un cálculo de la probabilidad logarítmica siempre que analiza un conjunto de efectos, lo que permite comparar diferentes análisis del mismo conjunto de datos. La formulación matemática básica es idéntica a la del modelo anterior

$$\frac{P}{1-p} = \frac{P_0}{1-p_0} \times \frac{P_i}{1-p_i} \times \frac{P_j}{1-p_j} \times \dots$$

donde P es la probabilidad de que una variante dada sea la utilizada en un contexto específico, P_0 es la probabilidad promedio sobre todos los contextos, y P_i , P_j , etc. los efectos de cada uno de los rasgos que definen dicho contexto. Sin embargo, el nuevo ejercicio que proporciona logra, aun bajo la hipótesis nula, para quien no es distinguible el efecto de rasgos conjuntos, que podemos saber, gracias al cambio de la probabilidad logarítmica y contrastando por separado el análisis de los dos y otro donde los dos factores aparezcan como uno y el mismo, si efectivamente un factor o un grupo de factores tiene un efecto estadísticamente significativo en la selección de la variante; tal determinación vendrá dada por la diferencia que arroje un text de χ^2 hecho sobre los resultados de la probabilidad logarítmica de los dos tipos de análisis.

Este es el punto principal del nuevo programa, pues gracias al uso sistemático y automático de esta probabilidad, el programa va com-

parando los resultados para llevar a cabo un análisis de regresión múltiple con una selección escalonada de factores. Cualquier variable lingüística puede, en teoría, resultar afectada por varios factores sociales y además por múltiples factores estructurales y diafásicos. Aunque se trate de un ejercicio complicado, todo investigador querrá saber cuál de los factores o grupo de factores es en realidad significativo en la selección de una variante. Este modelo, a través del programa VARBRUL 2S, nos proporciona esa información sin necesidad de analizar separadamente cada factor o grupo de ellos, tarea que además de compleja, no nos llevaría muy lejos, ya que estos factores no son estadísticamente independientes entre ellos.

¿Cómo se realiza este conjunto de operaciones? La selección escalonada de grupos de factores significativos se lleva a cabo mediante un proceso que comienza por calcular la probabilidad logarítmica de los datos en conjunto, sin tener en cuenta ningún factor. A este primer paso le siguen tantos análisis como factores entren en la consideración de la variable. Se compara después la probabilidad logarítmica de cada uno de ellos con la general, y previo cotejo de los resultados de la χ^2 correspondiente, el programa selecciona el grupo de factores que más patrocina la realización de la variante estudiada. El ejercicio se repite, pero ahora se trata de análisis bifactoriales, ya que se añade un nuevo factor al ya seleccionado con anterioridad. Si alguno de estos análisis arroja un aumento en cuanto a la probabilidad logarítmica superior al límite de entrada predeterminado (generalmente .05), entonces comienzan los análisis trifactoriales del resto de los factores; a los dos ya seleccionados se les van añadiendo, uno a uno, los restantes, y así sucesivamente.

El procedimiento puede ser corroborado con un análisis complementario reductor. Este examen incluye primero a todos los factores, calculando su probabilidad logarítmica; después por cada uno de los factores, se ve si ésta es significativamente menor si volvemos al análisis, esta vez omitiendo sólo el factor en cuestión. El menos significativo de todos resulta eliminado, y la operación continúa con los restantes factores. Al final de esta tarea se comprueba que los factores retenidos por el análisis reductor son (generalmente) los mismos que señala la selección escalonada de grupos de factores. De esta forma, Poplack comprueba que la elisión de *-s/* monomorfémica, por ejemplo, se debe a los siguientes factores:

palabra polisilábica	.65	monosilábica	.53
vocal siguiente	.63		
consonante sonora siguiente	.55	consonante sorda siguiente	.30
pausa siguiente	.54		
acento	.57	acento débil	.45
sexo masculino	.55	sexo femenino	.45
estilo espontáneo	.55	estilo informal	.50
		estilo formal	.45
escolaridad superior a cuatro años	.53	escolaridad inferior a cuatro años	.47
origen geográfico: este	.53	origen geográfico: oeste	.47

los factores que no fueron seleccionados por no ser relevantes al cumplimiento de la regla de elisión de *-/s/* en el dialecto puertorriqueño de Filadelfia son: edad y habilidad para hablar el inglés.

5. A pesar de todos los refinamientos técnicos que ha ido sufriendo el modelo probabilístico original, las bases teóricas de carácter lingüístico en las que se asienta aquél y los modelos logísticos más recientes son las mismas: presentar un nuevo modelo de competencia en el que quede inserto un componente probabilístico, componente que abarcará por igual factores lingüísticos y extralingüísticos. Pero esta nueva concepción choca con una fuerte y bien asentada tradición dentro de las corrientes lingüísticas de nuestro siglo, y concretamente, de las más actuales. Sankoff ha resumido con puntualidad ésta y otras actitudes contrarias a las sugerencias de que los conceptos probabilísticos merezcan alguna consideración en el estudio de la lengua. La «recepción hostil» de que habla, obedece a la convicción de que ni la casualidad (*randomness*) ni la indeterminación desempeñan ningún papel significativo en la estructura de la lengua, especialmente en lo que respecta a la competencia lingüística o ese saber lingüístico internalizado de todo hablante oyente.

El primer tipo de rechazo es de carácter histórico. Desde que Chomsky publicó sus *Syntactic Structures* en 1957 el punto de vista predominante en el mundo de la lingüística ha sido el generativo. El libro comienza por destruir dos modelos gramaticales basados en principios probabilísticos, aunque hartamente ingenuos. Aunque el ataque servía allí de motivación para crear una gramática generativa de estructura de frase y transformaciones, indirectamente avivó el descrédito de la probabilidad ante los ojos de muchos estudiosos. Más adelante, en su fundamental *Aspects of the theory of Syntax* de 1965,

Chomsky, dice expresamente que «aunque la probabilidad de una oración (tipo) sea clara y bien definida, es una noción totalmente inútil» (pág. 195).

Una segunda motivación para el rechazo tiene que ver con la metodología de la investigación gramatical. A pesar de que el lingüista dispone de una fuente extensa, directa y de acceso instantáneo para su estudio —los hablantes nativos de una lengua dada— en la práctica los análisis y las pruebas de aceptabilidad correspondientes son realizados sobre un solo sujeto, con frecuencia el mismo investigador. Aunque esto causa varios problemas, como la falta de reproductibilidad entre hablantes, se consideran como un pequeño precio que hay que pagar por todo lo que se logra saber sobre la estructura lingüística. Como un lingüista dado será internamente sistemático, no habrá variabilidad en sus datos y por lo tanto éstos no necesitarán ningún tratamiento estadístico. Por el contrario, la naturaleza invariable y discreta de un juicio individual sobre gramaticalidad se opone a la noción de medición cuantitativa de ensayos repetidos. Esta situación, sin embargo, es del todo insuficiente para rechazar los conceptos estadísticos pues es bien sabido y reconocido por todos, la existencia de variación inter-individual, dialectal e histórica existente en las lenguas. A pesar de esto, y de que gran parte de los estudios actuales sobre variación se ocupan precisamente de ello, muchos estudiosos de la variación se niegan a integrar la probabilidad en sus teorías y prefieren continuar produciendo análisis discretos y determinísticos que expliquen más o menos bien los fenómenos que estudian.

El tercero y más importante aspecto del anti-probabilismo tiene que ver, como asegura Sankoff, con la estructura de la lengua, especialmente como la entiende la lingüística moderna. No cabe duda de que las teorías lingüísticas actuales están formuladas en términos matemáticos; se trata de un tipo de matemáticas relacionada con la teoría de la programación lingüística, la teoría de los autómatas, la lógica, la teoría gráfica y otras disciplinas relacionadas que comúnmente se agrupan bajo el rótulo de «álgebra aplicada». El desarrollo histórico de estas disciplinas ha ganado con los avances en el estudio de la lengua, y no es raro ver que todas ellas y la lingüística compartan hasta las notaciones. En cambio, las extensiones probabilísticas de esas teorías son recientes, difíciles y no muy bien conocidas, quizá porque hasta ahora ha habido poca o escasa motivación empírica.

La ausencia de probabilidad y estadísticas en las corrientes más importantes de la lingüística actual se debe pues a la consideración de que los análisis estadísticos y los modelos probabilísticos estándar resultan inaplicables, por lo menos en sus formulaciones más usuales, a la estructura algebraica concebida por los lingüistas como estructura subyacente de la lengua. En vista de esto, la sugerencia aportada por Sankoff y realizada en los varios modelos ya vistos, es que el estudio estadístico de las lenguas se haga de una manera lingüísticamente aceptable, es decir, construyendo extensiones probabilísticas a los modelos lingüísticos algebraicos existentes.

No caben serias dudas de que la estructura algebraica de la lingüística formal incorpora de manera fundamental un mecanismo de selección que permite que una relativamente pequeña gramática genere un número teóricamente infinito de oraciones; sólo hay que mirar, entre otros casos también importantes, la lección que supone cumplir o no cumplir determinada regla sintáctica o fonológica en un contexto que lo permita. Pues bien, el hecho de que las estructuras gramaticales incorporan la «elección» como piedra fundamental indica que aceptan la probabilización de manera natural, hablando en términos matemáticos. Es precisamente el estudio de la variación lingüística la que ha motivado la creación de modelos probabilísticos y la que ha ofrecido justificación remuneradora para tales tareas; con ellos la estructura algebraica, originalmente categórica, discreta y abstracta, con la que se trata de explicar la estructura lingüística subyacente, se convierte en flexible, variable y útil como modelo para explicar la conducta lingüística real.

Aunque los modelos probabilísticos exceden con mucho los objetivos y los límites de la sociolingüística, es precisamente esta disciplina la que ha resultado más favorecida, al poder llevarse a feliz término la propuesta de Wienreich y sus colaboradores para que las lenguas fuesen comprendidas como estructuras ordenadamente heterogéneas; ya estamos en el camino de superar el reiterado interés por los «hablantes ideales» y propugnar el estudio directo de las comunidades lingüísticas.

HUMBERTO LÓPEZ MORALES
Instituto de Lingüística
Universidad de Puerto Rico

BIBLIOGRAFÍA

- Bailey, C. J., *Variation and linguistic theory*, Arlington, Center for Applied Linguistics, 1973.
- Bickerton, D., «Inherent variability and variable rules», *Foundations of Language*, 7, 1971, págs. 457-492.
- «Quantitative versus dynamic paradigms: the case of Montreal QUE», en *New ways of analyzing variation in English*, ed. por C. J. Bailey y R. W. Shuy, Washington, Georgetown University Press, 1973, págs. 23-43.
- Camp, D. de, «Toward a generative analysis of a post-creole speech continuum», en *Pidginization and creolization of languages*, ed. por D. Hymes, Cambridge, Cambridge University Press, 1971, págs. 349-370.
- Cedergren, H., *The interplay of social and linguistic factors in Panama*, tesis inédita, Ithaca, Cornell University, 1973.
- «On the nature of variable constraints», en *New ways of analyzing variation in English*, ed. por C. J. Pailey y R. W. Shuy, Washington, Georgetown University Press, 1973, págs. 13-22.
- «En torno a la variación de la S final de sílaba en Panamá: análisis cuantitativo», en *Corrientes actuales en la dialectología del Caribe hispánico*, ed. por H. López Morales, Río Piedras, Editorial Universitaria, 1978, págs. 35-50.
- y D. Sankoff, «Variable rules: performance as a statistical reflection competence», en *Language*, 50, 1974, págs. 333-355.
- «Nasals: A sociolinguistic study of change in progress», en *Nasálfest: Papers from a symposium on nasals and nasalization*, ed. por C. Ferguson, L. Hyman y J. Ohala, Stanford, Language Universals Project, 1975, págs. 67-80.
- Chomsky, N., *Syntactic structures*, La Haya, Mouton, 1957.
- *Aspects of the theory of syntax*, Cambridge /Mass./, MIT Press, 1965.
- Elliott, D., S. Legum y S. A. Thompson, «Syntactic variation as linguistic data», en *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, ed. por R. Binnick et al., University of Chicago Department of Linguistics, 1969, págs. 52-59.
- Fasold, R. W., «The Bailey wave model: A dynamic quantitative paradigm», en *Analyzing variation in English*, ed. por R. W. Fasold y R. W. Shuy, Washington, Georgetown University Press, 1975, págs. 27-58.
- «Language variation and linguistic competence», en *Linguistic variation: Models and methods*, ed. por D. Sankoff, New York, Academic Press, 1978, páginas 85-95.
- Kay, P., «Variable rules, community grammar, and linguistic change», en *Linguistic variation: Models and methods*, ed. por D. Sankoff, New York, Academic Press, 1978, págs. 71-83.
- y Ch. K. McDaniel, «On the logic of variable rules», reproducido por el Indiana Linguistics Club, Bloomington, 1977.

- Labov, W., «Contraction, deletion and inherent variability of the English copula», en *Language*, 45, 1969, págs. 715-762.
- López Morales, H., «Sociolingüística: nuevos enfoques metodológicos», en *Revista española de lingüística*, 7, 1977, págs. 17-36.
- Poplack, Sh., *Function and progress in a variable phonology*, tesis inédita, Filadelfia, University of Pennsylvania, 1979.
- Ross, John R., «The category squish: Endstation Hauptwort», en *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*, ed. por P. Peranteau et al., Chicago Linguistic Society, 1972, págs. 316-328.
- «A fake NP squish», en *New ways of analyzing variation in English*, ed. por C. J. Bailey y R. W. Shuy, Washington, Georgetown University Press, 1973, páginas 96-140.
- Rousseau, P., *Analyse de données binaires*, tesis inédita, Montreal: Université de Montréal, 1978.
- y D. Sankoff, «Advances in variable rule methodology», en *Linguistic variation: Models and methods*, ed. por D. Sankoff, New York, Academic Press, 1978, págs. 57-69.
- «A solution to the problem of grouping speakers», en *Linguistic variation: Models and methods*, ya citado, págs. 97-117.
- Sag, Ivan, «On the state of progress on progressives and statives», en *New ways of analyzing variation in English*, ya citado, págs. 83-95.
- Sankoff, D., *VARBRUL 2*, Programa y documentación inéditos, 1975.
- «Probability and linguistic variation», en *Synthese*, 37, 1978, págs 217-238.
- y P. Rousseau, «A method for assessing variable rule and implicational scale analyses of linguistic variation», en *Computers in the Humanities*, ed. por J. L. Mitchell, Edinburgh University Press, 1974, págs. 3-15.
- y W. Labov, «On the uses of variable rules», en *Language in Society*, 8, 1979, págs. 189-222.
- Sankoff, G., «Above and beyond phonology in variable rules», en *New ways of analyzing variation in English*, ya citado, págs. 44-61.
- Suppes, P., «Probabilistic grammars for natural languages», en *Synthese*, 22, 1970, págs. 95-116.