

Fundamentos de la lingüística forense

PROYECTO EDITORIAL CLAVES DE LA LINGÜÍSTICA

Director:
Juan Carlos Moreno Cabrera



Queda prohibida, salvo excepción prevista en la ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra sin contar con autorización de los titulares de la propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sigs. Código Penal). El Centro Español de Derechos Reprográficos (www.cedro.org) vela por el respeto de los citados derechos.

*Fundamentos
de la lingüística forense*

*Elena Garayzábal Heinze
Sheila Queralt Estévez
Mercedes Reigosa Riveiros*



Consulte nuestra página web: www.sintesis.com
En ella encontrará el catálogo completo y comentado

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previstos en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente, por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o por cualquier otro, sin la autorización previa por escrito de Editorial Síntesis, S. A.

© Elena Garayzábal Heinze
Sheila Queralt Estévez
Mercedes Reigosa Riveiros

Colaboradora: Irene Hidalgo de la Guía

© EDITORIAL SÍNTESIS, S. A.
Vallehermoso, 34. 28015 Madrid
Teléfono: 91 593 20 98
www.sintesis.com

ISBN: 978-84-9171-426-2
Depósito Legal: M. 32.622-2019

Impreso en España - Printed in Spain

Índice

1. Introducción a la lingüística forense	9
1.1. Lingüística forense en términos generales	10
1.2. Áreas de actuación de la lingüística forense	21
1.3. Marco teórico	25
1.3.1. Las diferentes ramas lingüísticas implicadas	25
1.3.2. La sociolingüística: la variación lingüística, estilo idiolectal y geolingüística	36
Preguntas de autoevaluación	41
2. Metodología de comparación en texto escrito	43
2.1. Valoración del material	43
2.2. Aproximación al análisis	45
2.2.1. Encargo	45
2.2.2. Material	46
2.2.3. Contexto de trabajo	47
2.3. Metodología de análisis	48
2.3.1. Patrón de análisis	48
2.3.2. Tipos de análisis	49
Preguntas de autoevaluación	56
3. Atribución de autoría	59
3.1. Introducción y definiciones de atribución de autoría	59
3.2. Metodología en el análisis forense de textos escritos	61

3.2.1. Variables de análisis en atribución de autoría	62
3.3. Herramientas del lingüista	65
3.4. Ámbitos de actuación	66
3.5. Casos relevantes en relación con la atribución de autoría	69
3.5.1. <i>El hijo del aviador y el carpintero (1932)</i>	69
3.5.2. <i>Las declaraciones de Evans (1950)</i>	70
3.5.3. <i>El asesino del policía (1952)</i>	71
3.5.4. <i>El caso del estudiante suicida (1992)</i>	73
3.5.5. <i>El caso Gilfoyle (1992)</i>	74
3.5.6. <i>El Unabomber (1995)</i>	74
3.5.7. <i>El caso de la reina de belleza infantil (1996)</i>	77
3.5.8. <i>El primer caso de análisis de SMS (2001)</i>	77
3.5.9. <i>El caso de la dependienta de supermercado (2005)</i> ..	79
3.5.10. <i>El canto del Cuco (2013)</i>	80
Preguntas de autoevaluación	81
4. Perfiles lingüísticos: la caracterización de autoría	83
4.1. ¿Cuándo se hace necesario recurrir a un perfil lingüístico?	84
4.1.1. <i>Bases teóricas en la elaboración de perfiles lingüísticos</i>	88
4.2. Metodología	95
4.3. Colaboración en la valoración de la amenaza durante la inves- tigación	98
4.4. Pasaporte vocal	101
4.5. Casos en los que fue necesario realizar un perfil lingüístico	103
4.5.1. <i>Operación Amerithrax</i>	104
4.5.2. <i>La línea del diablo</i>	106
4.5.3. <i>El perfil del autor de las cartas bomba en Estados</i> <i>Unidos</i>	106
4.5.4. <i>La farmacéutica de Olot</i>	107
4.5.5. <i>La tregua de ETA</i>	108
Preguntas de autoevaluación	108
5. Detección de plagio	111
5.1. Los tipos de plagio	112
5.1.1. <i>Copia de ideas</i>	112
5.1.2. <i>Plagio lingüístico</i>	113
5.2. El análisis en la detección de dependencia textual	114
5.2.1. <i>Análisis cualitativo</i>	115
5.2.2. <i>Análisis semiautomático</i>	116
5.2.3. <i>Los umbrales de similitud</i>	119
5.3. Directrices para detectar y evitar el plagio	120

Índice

5.4. Ámbitos de actuación	123
5.4.1. <i>Plagio musical</i>	123
5.4.2. <i>Plagio académico: cortar y pegar</i>	125
5.4.3. <i>Plagio literario</i>	131
5.4.4. <i>Plagio translingüe</i>	135
Preguntas de autoevaluación	139
6. Identificación de locutores	141
6.1. Bases de la fonética acústica	142
6.1.1. <i>¿Qué es y de qué se encarga la fonética acústica?</i> ...	142
6.1.2. <i>Terminología y métodos de la fonética acústica</i>	142
6.1.3. <i>La fonación, la cualidad de voz y los modos de fonación</i>	145
6.1.4. <i>Descripción de los sonidos del habla y sus correlatos</i>	
<i>acústicos</i>	151
6.1.5. <i>La cadena hablada</i>	164
6.2. La fonética acústica en el ámbito forense	167
6.2.1. <i>Aplicaciones de la fonética acústica en el ámbito</i>	
<i>forense</i>	168
6.2.2. <i>Metodología en comparación forense de habla</i>	171
6.3. Ruedas de reconocimiento de voz	177
6.4. Casos relevantes de fonética acústica forense	182
6.4.1. <i>Gürtel y sus distintas tramas</i>	183
6.4.2. <i>Operación Pokémon en Ferrol</i>	183
Preguntas de autoevaluación	184
7. La transcripción en el ámbito forense	187
7.1. La transcripción forense	188
7.2. Aspectos legales	193
7.3. Criterios de transcripción	195
7.4. Protocolo de actuación y buenas prácticas	200
7.4.1. <i>Respecto del perfil que ha de cumplir un transcriptor ..</i>	200
7.4.2. <i>Desde el punto de vista del propio transcriptor</i>	201
7.4.3. <i>Por lo que respecta a la propia investigación</i>	201
7.4.4. <i>Con relación a las fases del proceso de transcripción</i>	202
7.4.5. <i>En cuando a los datos relativos a la comunicación ..</i>	203
Preguntas de autoevaluación	204
8. El perito lingüista y el informe pericial	207
8.1. El perito forense	207
8.1.1. <i>Obligaciones y recusación de un perito</i>	211
8.2. El informe pericial y el dictamen	218

Fundamentos de la lingüística forense

8.2.1. Redacción del informe	221
8.2.2. Estructura del informe	226
8.2.3. Expresión de conclusiones en el informe forense	230
8.3. La defensa del peritaje en juicio	233
8.3.1. Sala de vistas	233
8.3.2. El perito en la sala de vistas	234
Preguntas de autoevaluación	237
Solucionario	239
Bibliografía	241

2

Metodología de comparación en texto escrito

Este capítulo se estructura en torno al protocolo de actuación del lingüista forense ante las muestras que deben ser objeto de análisis en un proceso de comparación de muestras, puesto que es el escenario más frecuente en lingüística forense. En particular, trata sobre cómo debe valorarse el material para determinar la existencia de caso lingüístico o no, cómo se determinan los procedimientos metodológicos más adecuados para cada caso teniendo en cuenta el encargo, el material de estudio y el contexto más externo, como puede ser el tiempo del que se dispone para la ejecución de la comparación y, finalmente, los distintos tipos de análisis que se pueden implementar teniendo en cuenta también los factores anteriores.

2.1. Valoración del material

El primer paso en cualquier análisis de comparación forense es valorar el material objeto de estudio, tanto por lo que concierne a las muestras cuya autoría desconocemos (muestras dubitadas) como a las muestras del sospechoso o de los distintos sospechosos (muestras indubitadas). Para valorar ambos tipos de muestras se tienen en cuenta dos condiciones imprescindibles: longitud y calidad (Queralt 2014: 37):

1. Que las muestras tengan un número de palabras o duración suficiente. Por un lado, en la comparación de muestras escritas, a pesar de que no existe un mínimo establecido sobre cuántas palabras debe tener un texto (Stamatatos, 2009), es cierto que los estudios cualitativos no pueden trabajar con muestras de menos de 150-200 palabras y los estudios cuantitativos necesitan un míni-

mo de 200, 500 y 1 000 palabras, y ciertamente la precisión de sus resultados decrece significativamente cuando la longitud del texto es inferior a las 1 000 (Hirst y Feiguina, 2007). En el caso de la comparación de muestras de audio, ocurre algo similar: no existe una duración mínima preestablecida de la grabación por analizar, en lo que respecta a la aproximación clásica, aunque sí en el caso de caso del análisis automático, dado que los programas necesitan una mínima cantidad de habla neta del locutor tanto para construir un modelo (algo más de 15 segundos) como para introducir las muestras de comparación que deberán durar, al menos, entre 40 y 60 segundos. En lo que respecta al análisis clásico, la insuficiencia cuantitativa quedaría determinada por el sentido común, teniendo en cuenta que unos breves segundos de muestra no permiten llevar a cabo una posterior comparación de habla. Lo ideal para realizar esta es disponer de alrededor de 10 frases completas extraídas de la muestra dubitada que presenten calidad suficiente para su co-tejo, y eso puede suponer que se disponga de al menos de más de 1 minuto de grabación.

2. Que las muestras sean de calidad. Tanto en las muestras escritas como en las muestras orales, se deben ver reflejadas variables que aporten información sobre algunos de los rasgos sociolingüísticos del individuo, como marcas de su procedencia geográfica, contacto de lenguas, nivel educativo, profesión, sexo, etc. En la comparación forense de muestras orales, esta calidad también se interpreta en términos relativos a cuestiones vinculadas al proceso de grabación de dichas muestras; no solo con el dispositivo usado en la grabación, sino también con el contexto en el que esta se realiza, así como con la cantidad de hablantes que participan en ella o el nivel de sonidos no deseados; es decir, ruidos registrados que dificultan o impiden el análisis acústico (véase capítulo 6).

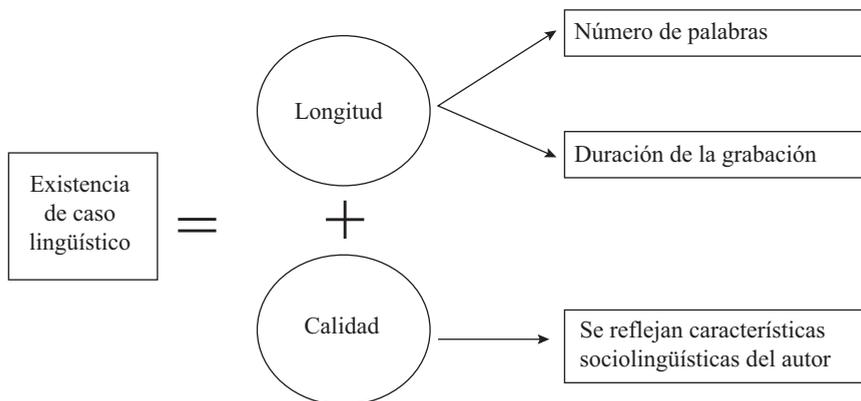


Figura 2.1. Condiciones para determinar la existencia de caso lingüístico.

Para que un análisis de comparación forense pueda llevarse a cabo por parte del lingüista forense ambas condiciones deben cumplirse. En el caso contrario, el experto deberá rechazar el encargo argumentando que el material no es suficiente por su longitud o por su calidad.

2.2. Aproximación al análisis

Una vez que se ha valorado el material y se ha determinado que existe caso lingüístico, puesto que se dispone de suficiente material y, además, este ofrece pistas sobre las características del autor, el lingüista forense debe preguntarse qué teorías son aplicables para el análisis y qué técnicas son más adecuadas para el caso.

El marco teórico aplicable a cada caso va a depender del tipo de encargo que se atienda. No hay un protocolo de teorías aplicables, ya que la casuística forense puede ser muy diversa, pero sí que suelen seguirse unas teorías y premisas principales para cada ámbito de aplicación. Así, por ejemplo, en el ámbito de la comparación forense de muestras orales o escritas destacan las teorías sociolingüísticas de la variación y el cambio lingüístico, así como el concepto de estilo idiolectal e incluso para determinar la posibilidad de que una conversación no haya sido espontánea, se aplicarán teorías como la teoría de la (des)cortesía de Brown y Levinson (1987). Como se puede observar, es imprescindible que el perito sea un buen lingüista, un buen investigador de la lengua, que sepa qué buscar y cómo hacerlo, que conozca las teorías del campo aplicables para poder determinar cuáles utilizar y qué métodos aplicar, así como conocer las teorías complementarias que resulten pertinentes en cada caso, que justifiquen el procedimiento de análisis implementado por el experto en cada asunto.

Para determinar la metodología que se va a aplicar hay tres puntos clave que deben tenerse en cuenta cuando un perito se aproxima a la muestra en un contexto forense real: el encargo, el material y el contexto de trabajo.

2.2.1. Encargo

El objetivo del encargo debe ser claro y preciso desde el inicio. El propósito del análisis siempre debe formularse de la forma más objetiva posible y en la que no se pueda entrever una orientación partidista en el análisis. Por ejemplo, en los casos de comparación, el objetivo nunca debe ser determinar si X es el autor de Y o si Z ha plagiado a P, sino que el objetivo debe ser realizar una comparación para determinar la probabilidad de que X sea el mismo autor de Y o realizar un análisis comparativo para determinar la probabilidad de que Z y P se hayan producido de forma independiente.

2.2.2. Material

El material objeto de estudio también se debe acotar en el inicio del encargo. En el campo de las ciencias forenses, el material se clasifica en términos de material dubitado e indubitado.

Material dubitado: se denomina *dubitado* al elemento de prueba sobre cuya autenticidad se tienen dudas, lo que puede significar que se desconoce el autor, el origen u otros elementos. En lingüística forense, el material dubitado se corresponde, por ejemplo, con grabaciones anónimas o textos o marcas comerciales que han podido ser plagiadas. Será siempre el primer material objeto de análisis. Hay que recordar, como se ha dicho más arriba, que en muchas ocasiones solo se dispone de material dubitado. El hecho de afrontar en primer lugar el análisis del texto dubitado evita que se contamine de las características de los indubitados y que, por tanto, se haga una aproximación de forma mucho más objetiva.

Material indubitado: se denomina *indubitado* al elemento de prueba sobre cuya autenticidad no hay dudas. En lingüística forense, el material indubitado corresponde, por ejemplo, a las muestras orales o escritas del sospechoso, al texto o la marca comercial original en un posible caso de plagio.

En el caso de la comparación, el material de cotejo debe ser comparable, es decir, material en el que se puedan establecer correspondencias en la naturaleza de las muestras entre el corpus dubitado y el corpus indubitado. Para valorar si las muestras son comparables, el experto lingüista suele regirse según los elementos del proceso comunicativo:

Cuadro 2.1. Elementos para determinar si las muestras son comparables

<i>Receptor</i>	El receptor de los mensajes indubitados debe tener unas características lo más similares posibles al receptor de los mensajes dubitados.
<i>Mensaje</i>	Se debe tener en cuenta el contenido del mensaje por lo que concierne a los temas tratados.
<i>Código</i>	Es recomendable que las muestras estén formuladas en la misma lengua.
<i>Canal</i>	Es importante que las muestras dubitadas y las indubitadas pertenezcan, en el caso de la comparación forense de textos, a géneros textuales lo más similares posibles. En el caso de comparación forense de habla, sobre todo si se recurre al uso del sistema de reconocimiento automático de locutores, las grabaciones deben haberse producido por el mismo canal (p. ej., GSM, micrófono, etc.).
<i>Contexto</i>	El contexto es un factor extremadamente difícil de controlar y reproducir. Por este motivo, es un factor cuya similitud es recomendable, pero no exigible. Cuando se hace referencia a la comparación de muestras de habla, es aconsejable realizar las muestras indubitadas con una selección de frases idénticas a las emitidas en las muestras de voz dubitada.

En primer lugar, el receptor. El grado de familiaridad entre el emisor y el receptor de los mensajes indubitados debe ser, en la medida de lo posible, lo más similar posible al grado de familiaridad entre el emisor y el receptor de los mensajes dubitados. Por ejemplo, si la llamada anónima es a una centralita del 112 y el emisor se expresa con un estilo formal, es preferible que las muestras orales indubitadas sean a agentes externos y no familiares. Además, también se puede considerar si el mensaje va destinado a un público amplio o a un individuo en particular.

En segundo lugar, el mensaje. Se debe tener en cuenta el contenido del mensaje por lo que concierne a los temas tratados. La similitud entre las muestras de este elemento puede facilitar el análisis, pero en ningún caso una falta de correspondencia en la temática de los mensajes imposibilita la comparación.

En tercer lugar, el código. Es recomendable que las muestras estén formuladas en la misma lengua, aunque se ha demostrado que también es posible realizar comparaciones entre muestras de distintas lenguas, como en Marquina (2016). Es altamente recomendable que el perito conozca ambos códigos o que se forme un equipo pericial en que existan expertos de ambas lenguas.

En cuarto lugar, el canal. El soporte físico es un factor clave para determinar la viabilidad del estudio, en especial en algunos ámbitos de comparación, como el de muestras orales. En el caso de las muestras orales se tiene en cuenta si estas son llamadas a través de un teléfono móvil, audios registrados mediante una grabadora o mensajes de voz. En el caso de muestras escritas, se tiene en cuenta el género textual de las muestras, ya sean escritos de mensajería instantánea, correos electrónicos, entradas de blog o comentarios en YouTube. No obstante, en todos los casos es importante que las muestras dubitadas y las indubitadas pertenezcan a géneros textuales lo más similares posibles. Dos muestras de géneros totalmente opuestos pueden imposibilitar el análisis o limitarlo de forma significativa.

En quinto lugar, el contexto. El contexto es un factor que puede repercutir en el análisis, pero que difícilmente se puede exigir como un elemento que deba tenerse en cuenta para determinar si el material es comparable, dado que, en el contexto forense, las producciones dubitadas suelen surgir en contextos con una alta carga emocional (amenazas, extorsiones, quejas entre gritos y llantos) que no suele producirse en las muestras indubitadas.

Para concluir, se debe destacar que serán en último lugar el conocimiento y la experiencia del perito lingüista forense los que determinarán, en caso de muestras dispares, la viabilidad y las limitaciones de la comparación.

2.2.3. Contexto de trabajo

Es importante determinar en qué contexto interviene el lingüista, teniendo en cuenta si se trata de un contexto de investigación o de un contexto probatorio. En el primer caso, el perito lingüista interviene durante la investigación de un caso para aportar,

generalmente, rasgos sociolingüísticos del autor de las muestras dubitadas como son el origen dialectal, el sexo o el nivel educativo, con el fin de intentar estrechar el círculo de sospechosos. En este contexto, el perito lingüista también puede desarrollar otras tareas, como pueden ser la limpieza y la transcripción de grabaciones (Cicres y Gavaldà, 2014: 62). En el contexto probatorio, el perito aporta pruebas lingüísticas durante el proceso judicial, generalmente en el ámbito de la comparación de muestras orales o escritas para determinar la autoría o la existencia de plagio.

También es imprescindible conocer el tiempo del que se dispone para hacer la comparación. No es extraño, en el contexto forense, disponer de tan solo unos pocos días, y, por tanto, es necesario acotar qué material va a poder analizarse y qué técnicas o tipos de análisis van a poder incorporarse. En el caso de contextos probatorios, muchos casos deben rechazarse a pesar de contar con un material de estudio adecuado, porque es imposible llevarlos a cabo en el plazo marcado. Además, no se debe tener en cuenta solo el tiempo de análisis, sino también el tiempo que va a requerir confeccionar el dictamen pericial. En el contexto de investigación, sin embargo, el tiempo de análisis suele ser más limitado y el perito debe ajustarse más a los tiempos. Además, en el proceso de investigación, el tiempo de redacción del informe suele ser más breve, ya que en muchos casos solo se debe redactar un informe breve o formular verbalmente las conclusiones del estudio a los agentes policiales o al cliente.

2.3. Metodología de análisis

A continuación, se expone el patrón de análisis que se debe seguir en toda pericial lingüística a la vez que se exponen los distintos métodos de análisis posibles teniendo en cuenta la casuística forense.

2.3.1. Patrón de análisis

El análisis lingüístico forense debe seguir el patrón común de cualquier estudio empírico experimental en el que se comienza con el planteamiento del problema, la fijación de unos objetivos, la delimitación de la metodología y los procedimientos que se van a aplicar; la obtención de los datos, el análisis de estos, su interpretación y, finalmente, las conclusiones.

En el caso del análisis lingüístico forense, el patrón se transforma en:

1. La valoración de la existencia de caso lingüístico.
2. La fijación del objetivo del estudio.
3. La delimitación de las teorías y los métodos aplicables más adecuados para el caso.
4. La valoración del material comparable dentro del material disponible.

5. El análisis lingüístico de los datos.
6. La interpretación de los resultados.
7. La formulación de las conclusiones de la pericial.

Es muy importante aplicar siempre un protocolo en el que se explicita la teoría o teorías que se vayan a aplicar, la verificación o refutación de las hipótesis formuladas en función del objetivo fijado y la determinación de las distintas variables que permitirán llegar a una conclusión.

2.3.2. Tipos de análisis

El análisis lingüístico de comparación de muestras se aborda desde estudios cualitativos, ofreciendo descripciones de las variables lingüísticas analizadas o desde estudios cuantitativos mediante la incorporación de técnicas estadísticas que ofrecen una organización de los datos y valores numéricos de significación de los resultados. Este apartado se ampliará en los capítulos correspondientes a cada campo de actuación.

Los análisis cualitativos se centran en el estudio descriptivo de las variables que caracterizan una muestra lingüística y suelen implementarse cuando se dispone de escasa cantidad de muestras, muestras de poca longitud o cuando no se dispone de suficientes realizaciones de la variable, pero, no obstante, resultan estructuras marcadas y de especial relevancia para describir el estilo idiolectal del autor. Si, por ejemplo, en el conjunto de muestras dubitadas y en el conjunto de las muestras del sospechoso aparece “lunnes” en lugar de “lunes”, resultaría de especial relevancia dicha idiosincrasia. Lo mismo ocurre en el caso de muestras de voz en una elisión de, por ejemplo, “gacias” por “gracias” cuando la simplificación del grupo consonántico se produce una única vez en ambos conjuntos de muestras; no sería un rasgo distintivo del hablante que se deba tener en cuenta en la valoración si no se produce la repetición de esa característica en más ocasiones.

Los análisis cuantitativos pueden dividirse de forma general entre análisis descriptivos e inferenciales. Por un lado, los análisis cuantitativos descriptivos permiten organizar los datos, proporcionar métodos de presentación y alcanzar conclusiones válidas. Generalmente, se ofrecen distribuciones en tablas y gráficos, por puntuación o por porcentaje, para establecer las diferencias o similitudes entre diferentes grupos de datos.

A continuación, se ofrece un ejemplo ficticio de análisis de la variable “número de oraciones por párrafo” en el que se analiza el número de oraciones de 10 párrafos que integran un texto dubitado, el número de oraciones de 10 párrafos que integran un texto indubitado del sospechoso 1 (Indubitado 1) y el número de oraciones de 10 párrafos que integran un texto indubitado del sospechoso 2 (Indubitado 2). Todos los análisis se han hecho con uno de los programas estadísticos informáticos más utilizados en las ciencias sociales y aplicadas: Statistical Package for the Social Sciences (SPSS).

Uno de los métodos descriptivos más utilizados es el uso de porcentajes de realización de la variable. Más abajo se exponen tres tablas correspondientes a cada uno de los autores, dubitado, indubitado 1 e indubitado 2, respectivamente. En la primera columna de cada tabla se muestra el número de oraciones por párrafo; en la segunda columna, la frecuencia, es decir, cuántas veces se ha realizado ese número de oraciones por párrafo; y, finalmente, en la tercera columna se presenta el porcentaje que representa ese número de oraciones por párrafo dentro del total de ese escritor. De este modo, el texto dubitado (cuadro 2.2) se compone por párrafos de 3, 4, 5, 6 y 7 oraciones, según la columna 1. En la columna 2 se puede leer que aparecen 3 párrafos de 3 oraciones, 2 de 4 oraciones, 3 de 5 oraciones, 1 de 6 oraciones y 1 de 7 oraciones. Y, por último, en la columna 3, observamos que en un 30 % de las ocasiones el autor utiliza párrafos de 3 o 5 oraciones, en el 20 % de los casos de 4 oraciones y en un 10 % de las ocasiones de 6 o 7 oraciones.

El análisis cuantitativo descriptivo también permite representar estos mismos datos de forma visual para facilitar su comprensión. A continuación, en las figuras 2.2, 2.3 y 2.4, se muestra la representación visual de los datos expuestos en sus correspondientes cuadros.

Cuadro 2.2. Número de oraciones por párrafo dubitado

	<i>Frecuencia</i>	<i>Porcentaje</i>	
<i>Válido</i>	3,00	3	30,0
	4,00	2	20,0
	5,00	3	30,0
	6,00	1	10,0
	7,00	1	10,0
	Total	10	100,0

Cuadro 2.3. Número de oraciones por párrafo indubitado 1

	<i>Frecuencia</i>	<i>Porcentaje</i>	
<i>Válido</i>	1,00	1	10,0
	3,00	2	20,0
	4,00	2	20,0
	5,00	1	10,0
	6,00	2	20,0
	7,00	2	20,0
	Total	10	100,0

Cuadro 2.4. Número de oraciones por párrafo indubitado 2

	Frecuencia	Porcentaje	
Válido	5,00	1	10,0
	6,00	2	20,0
	7,00	2	20,0
	8,00	1	10,0
	9,00	2	20,0
	10,00	1	10,0
	11,00	1	10,0
	Total	10	100,0

También se pueden utilizar medidas de tendencia central, como son la moda, la mediana y la media, para determinar cuál es la frecuencia de uso de una variable particular de un autor y si se corresponde con la frecuencia de uso del autor de los textos dubitados:

- *Moda*: la puntuación con mayor frecuencia en una distribución de datos, esto es, el dato que más veces se repite.
- *Mediana*: la puntuación situada en el centro de la distribución de una serie de datos ordenados.
- *Media*: distribución de las puntuaciones en una escala cuyo promedio es el resultado de sumar todas las puntuaciones y dividir por el número de ellas.
- *Desviación estándar*: indica la dispersión de los datos con relación a la media, es decir, cuanto mayor sea la desviación estándar, mayor será la dispersión de los datos. Se suele utilizar para establecer un valor de referencia que pueda ayudar a determinar una variación respecto de una situación estándar.

A pesar de que una persona tenga una frecuencia de uso (media) de una variable, también presenta variación en esa variable. Por ejemplo, un escritor puede formar oraciones de 5 palabras, de 7 palabras o de 6 palabras, pero siempre dentro de una horquilla más o menos acotada. Las medidas de tendencia típica, como la desviación estándar, nos permiten medir esa horquilla o rango de frecuencias de la variable que puede realizar un autor. La desviación estándar indica de qué forma los valores de los datos difieren de la media; es decir, se centra en el estudio de la variabilidad de los resultados alrededor de la media. Por lo tanto, cuanto más alta es la desviación estándar, más variación respecto del punto central de la distribución; en consecuencia, más alejada de la norma y, por ello, más significativa; cuanto más baja es la desviación estándar, más cercana está la distribución al punto central (cuadro 2.5).

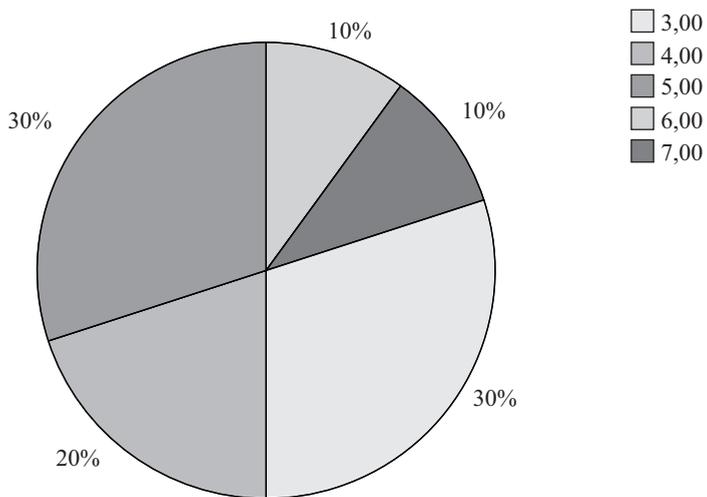


Figura 2.2. Número de oraciones por párrafo dubitado.

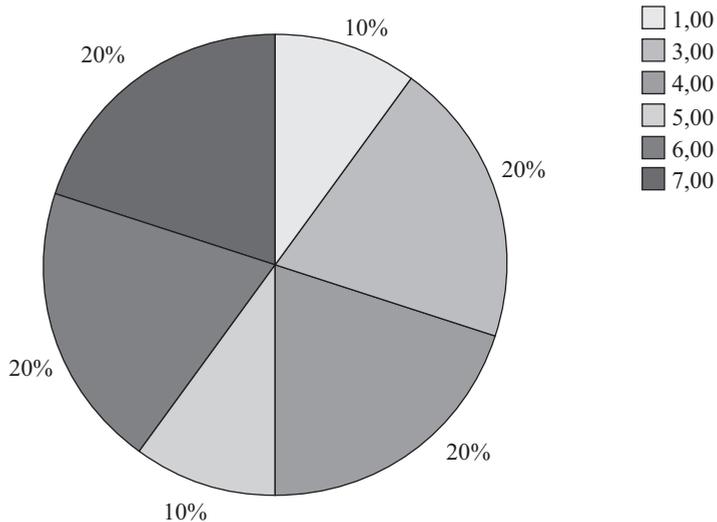


Figura 2.3. Número de oraciones por párrafo indubitado 1.

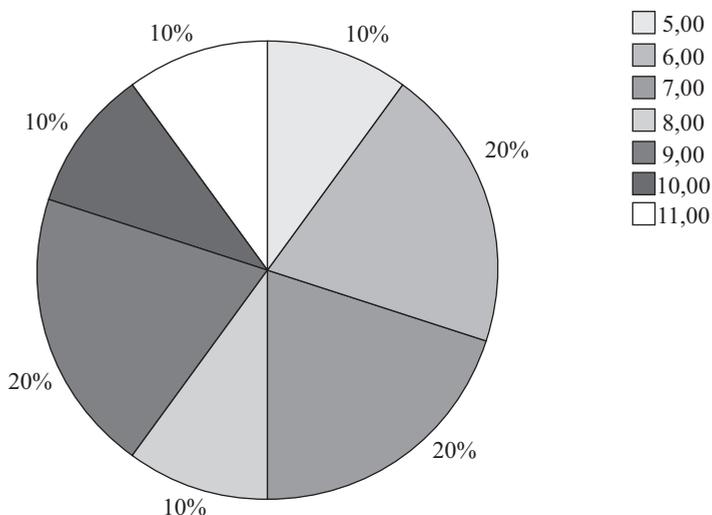


Figura 2.4. Número de oraciones por párrafo indubitado 2.

Cuadro 2.5. Datos descriptivos

	<i>N</i>	<i>Moda</i>	<i>Mediana</i>	<i>Media</i>	<i>Desviación estándar</i>
Dubitado	10	5	4,5000	4,5000	1,35401
Indubitado 1	10	6	4,5000	4,6000	1,95505
Indubitado 2	10	6	7,5000	7,8000	1,93218
Total	30	5	5,5000	5,6333	2,31164

Por lo que concierne al análisis estadístico cuantitativo inferencial, en lingüística forense se parte del planteamiento de dos hipótesis:

- *Hipótesis nula*: indica que no hay diferencias entre las muestras dubitadas y las muestras indubitadas.
- *Hipótesis alternativa*: se acumula evidencia suficiente para rechazar la hipótesis nula y, por tanto, se puede determinar que existen diferencias estadísticamente significativas entre ambos conjuntos de muestras.

Entre las técnicas cuantitativas más utilizadas en lingüística forense destacan las pruebas de significación estadística como el ji-cuadrado, el análisis de la varianza (ANOVA) y el análisis discriminante. El objetivo de las pruebas de significación es

demostrar si las diferencias observadas entre los conjuntos dubitado e indubitado pueden atribuírsele a la variación producida por el propio muestreo o porque realmente se produce un contexto de variación de una de las variables independientes cuyo efecto en la variable dependiente (conjunto) está siendo investigado.

A continuación, en el cuadro 2.6, y siguiendo con el ejemplo anterior de la variable “número de oraciones por párrafo”, se muestran los resultados del test ANOVA.

Cuadro 2.6. Resultado ANOVA

<i>Número de oraciones por párrafo</i>					
	<i>Suma de cuadrados</i>	<i>gl</i>	<i>Media cuadrática</i>	<i>F</i>	<i>Sig.</i>
Entre grupos	70,467	2	35,233	11,258	0,000
Dentro de grupos	84,500	27	3,130		
Total	154,967	29			

El p-valor es una medida que indica la probabilidad de que la diferencia observada se deba o no al azar. El resultado del p-valor (Sig. en la tabla) del test ANOVA menor a 0,05 indica que existen diferencias estadísticamente significativas entre los grupos. Para conocer entre qué grupos existen diferencias, es decir, qué textos son similares y qué textos son distintos según la variable “número de oraciones por párrafo”, se aplica una prueba *post-hoc*. Estas pruebas permiten determinar qué medias difieren. En este ejemplo se utiliza la prueba de Tukey, una de las pruebas más frecuentes para este tipo de datos.

Cuadro 2.7. Resultado prueba Tukey

<i>Comparaciones múltiples</i>			
Variable dependiente: Número de oraciones por párrafo			
<i>HSD Tukey</i>			
<i>(I) Autor</i>	<i>(J) Autor</i>		<i>Sig.</i>
Dubitado	Indubitado 1		0,991
	Indubitado 2		0,001

* La diferencia de medias es significativa en el nivel 0,05.