UNIVERSIDAD DE CHILE
FACULTAD DE FILOSOFÍA Y HUMANIDADES
ESCUELA DE POSTGRADO

UN ESTUDIO COMPARATIVO DE LA FRECUENCIA DE OCURRENCIA DE LAS EXPRESIONES EVALUATIVAS EN TEXTOS EXPOSITIVOS CIENTÍFICOS EN INGLÉS

Tesis para optar al grado de Magíster en Lingüística mención Lengua Inglesa

ESTUDIANTE: ENRIQUE PRIETO E.

PROFESOR GUÍA: CARLOS ZENTENO B.

SANTIAGO DE CHILE
2012

What more powerful form of study of mankind could there be than to read our own instruction book?

Francis S. Collins,
American geneticist

ÍNDICE

1. INTRODUCCIÓN

En las últimas décadas, particularmente en el ámbito del análisis del discurso y de la lingüística textual, se han propuesto nuevas aproximaciones teóricas y modelos descriptivos aplicables al estudio de los recursos formales de las lenguas empleados en la expresión de los significados afectivos y evaluativos. Estas investigaciones han tomado en cuenta diferentes ejes conceptuales: 'afecto' (Ochs y Schieffelin 1989), 'posición del hablante' ('stance') (Biber et al. 1999, Jaffe 2009), 'evaluación' (Hunston 1994, Hunston y Thompson 2001) y 'actitud' (Martin 2000a, Martin y White 2005), principalmente.

En el dominio de los estudios de la lengua inglesa, los significados afectivos y evaluativos se manifiestan en distintos 'modos discursivos', o 'tipos de discurso' (cf. Smith 1997, Ciapuscio 1994) mediante diferentes aproximaciones teórico-analíticas. Al respecto, han sido influyentes las propuestas entregadas por el análisis de la conversación casual (Eggins y Slade 1997), la teoría de la valoración (Martin 2000a, Macken-Horarik y Martin 2003, Martin y Rose 2003, Martin y White 2005) y el modelo sistémico-funcional de la evaluación (Hunston y Sinclair 2001), principalmente. En general, estos modelos descriptivos se inscriben en la lingüística sistémico-funcional (Halliday 1985, 1994, Halliday y Mathiessen 2004). La mayor parte de ellos permiten una explicación sistemática de los significados afectivos y evaluativos mediante un conjunto de categorías descriptivas válidas y aplicables al estudio de otros modos y géneros discursivos. No obstante, solamente un número menor de investigaciones han dado cuenta del funcionamiento sistemático de los recursos sintácticos que formalizan los significados referidos en textos orales o escritos en inglés. Al respecto, deben mencionarse las propuestas de Channell (2001), Hunston (1994), y Hunston y Sinclair (2001).

De la cuenta anterior, no deberían excluirse otras propuestas acerca de la descripción de los significados afectivos y evaluativos. Al respecto, se destaca el trabajo seminal de Labov (1972)[1] del componente evaluativo en el discurso narrativo oral. En cuanto al funcionamiento del componente evaluativo en otras modalidades discursivas, pueden mencionarse los estudios de Bolívar (1997, 2005), primero, de la negociación de la evaluación en editoriales periodísticos y, luego, sobre el papel de la evaluación en el diálogo y confrontación en el medio político venezolano; Harvey (2005), quien estudia, en una aproximación sociopragmática, las expresiones evaluativas en la ciencia como discurso; Johansson (2001) y su investigación, inserta en el análisis crítico del discurso, de la relación entre discurso e ideología en editoriales de medios de prensa escrita acerca del 'Informe de la Comisión Verdad y Reconciliación' a través del análisis de las clases léxicas adjetivos y sustantivos; Kerbrat-Orecchioni (1993), con su propuesta de los significados valorativos de las clases léxicas adjetivos, adverbios y sustantivos en el francés; Shiro (1999, 2004), con su examen de la relación entre evaluación y habilidades narrativas de adultos y de niños venezolanos; finalmente, Zenteno (2002-2003)[2], quién da cuenta del funcionamiento de los significados valorativos en el discurso argumentativo presente en editoriales de medios de prensa escrita.

[1] Referido a este punto, Shiro (1999) afirma que "…el discurso narrativo es un medio fundamental para expresar la subjetividad. La mayoría de los estudios sobre narrativa enfatizan la característica dual de este género. Labov (1972) sostiene que las narraciones tienen dos funciones básicas: la referencial y la expresiva. La función referencial sirve para reportar la secuencia de eventos en la que se basa el relato. La función expresiva, en cambio, sirve para construir la perspectiva narrativa". (1999:1)

[2] Al respecto, Zenteno (2002-2003) señala que en esta clase textual "…el especialista examina tal estado de cosas para proceder a expresar una serie de opiniones que contienen juicios apreciativos tanto de la problemática en general como de aspectos específicos de ella. Mediante tales opiniones, el enunciador pone de manifiesto su particular evaluación del problema, sustentándola ya sea sobre la base de sus conocimientos de experto o, como parece ser el caso, en evidencia empírica derivada de su propia investigación". (2002-2003:307)

En nuestra opinión, el componente evaluativo en textos científicos no parece constituirse aún en un objeto de estudio sistemático en los dominios del análisis del discurso o de la pragmática. Dicha ausencia originó nuestro interés por abordar su estudio como una primera aproximación que, fundada en los modelos descriptivos de Hunston y Sinclair (2001), Martin y White (2005) pudieran establecerse ciertas bases sistemáticas para otras investigaciones referidas al tema. En concordancia con lo anterior, los objetivos y características del presente trabajo implican una revisión y, en parte, una consecuente modificación de las categorías descriptivas de tanto los significados afectivos y evaluativos como de su formalización sintáctica en publicaciones científicas académicas y en medios de divulgación de esta actividad al público en general, todos los cuales son representativos del discurso expositivo en inglés.

Nuestra investigación ha consistido en un estudio comparativo de la frecuencia de ocurrencia de tanto los 'patrones sintácticos evaluativos' (Hunston y Sinclair 2001) como de los 'significados valorativos' (Martin y White 2005) i.e., significados semántico-pragmáticos evaluadores empleados en textos expositivos científicos escritos en inglés. En este tipo textual[3], se incluyen, primero, artículos publicados en revistas[4] científicas; segundo, artículos periodísticos de investigación científica publicados en revistas destinadas al público en general y tercero, editoriales publicados en números extraordinarios de revistas de investigación científica.

[3] Ciapuscio (1994) propone una distinción entre los términos 'tipo textual' y 'clase textual'. El primero referido a "...clasificaciones empíricas [...] realizadas por los miembros de una comunidad lingüística" y el segundo definido como "...una categoría ligada a una teoría para la clasificación [...] de textos. Al respecto, Ciapuscio (1994) afirma que "...por lo tanto, los hablantes de una comunidad tienen un saber sobre clases textuales o un saber sobre estructuras textuales globales pero no un saber sobre tipos textuales" (1994:25). Smith (1997), por otra parte, distingue entre 'modo discursivo' y 'género discursivo', respectivamente.

[4] En el transcurso de la presente investigación, se utilizará el término 'revistas' como traducción equivalente del término en inglés 'journals'.

En su totalidad, los textos fueron seleccionados entre los años 2000 y 2008, todos los cuales son versiones electrónicas de las publicaciones 'Bulletin of the World Health Organization', 'Current Science', 'Discover Magazine', 'Gene', 'Genetics', 'Genome Research', 'Human Genome Collection', 'National Geographic News', 'Nature', 'Proceedings of the National Academy of Sciences of the United States of America', 'Molecular Biology and Evolution', 'Science', 'Science News', 'Scientific American' y 'Time Magazine'. El corpus recopilado presenta la discusión del denominado 'Proyecto Genoma Humano (PGH)'. El objetivo fundamental de este programa internacional de investigación científica es determinar la secuencia de pares de bases químicas que componen el ADN e identificar y cartografiar los aproximadamente 25,000 genes del genoma humano. El PGH ha sido desarrollado, desde el año 1990, por el gobierno de los Estados Unidos de Norteamérica bajo la dirección del biólogo estadounidense James Watson con un plazo de realización de 15 años. Debido a la amplia colaboración internacional y los avances científicos, en el campo de la biología y la tecnología computacional, la secuencia final del genoma humano fue finalizada en el año 2003.

Las distintas etapas del estudio realizado se presentan aquí. Con respecto a la organización formal se establecen diez secciones. La sección uno es la introducción de la presente investigación. La sección dos presenta la descripción de los objetivos generales y de los objetivos específicos. En la sección tres se introduce una revisión de la literatura especializada. La sección cuatro presenta la hipótesis central y la hipótesis nula. En la sección cinco se exponen las preguntas de investigación. La sección seis describe el método de investigación, aquí se sintetiza la descripción del corpus, los criterios de selección del objeto de estudio y los procedimientos de análisis. En la sección siete se presentan, ilustran y discuten los resultados obtenidos del análisis. En la sección ocho se establecen las

conclusiones e implicancias de la investigación. En la sección nueve se listan las referencias bibliográficas y, finalmente, la sección diez compendia los anexos.

2. OBJETIVOS

2.1 Objetivos generales

2.1.1 Realizar un estudio comparativo de los patrones sintácticos empleados en la expresión de significados evaluativos (Hunston y Sinclair 2001) presentes en textos expositivos científicos escritos en inglés.

2.1.2 Realizar un estudio comparativo de las categorías de significados evaluativos presentes en textos expositivos científicos escritos en inglés.

2.2 Objetivos específicos

2.2.1 Identificar y categorizar los ‘patrones sintácticos evaluativos’ presentes en textos expositivos científicos escritos en inglés.

2.2.2 Determinar la frecuencia de ocurrencia de los patrones sintácticos evaluativos presentes en textos expositivos científicos escritos en inglés.

2.2.3 Identificar y categorizar los ‘significados valorativos’ expresados en textos expositivos científicos escritos en inglés.

2.2.4 Determinar la frecuencia de ocurrencia de los significados valorativos expresados en textos expositivos científicos escritos en inglés.

2.2.5 Comparar las modalidades de funcionamiento de los patrones sintácticos evaluativos y los significados valorativos en los textos bajo estudio con el fin de establecer un modelo

descriptivo preliminar del componente evaluativo del discurso expositivo científico en inglés.

3. MARCO TEÓRICO

3.1 Algunas aproximaciones al concepto 'evaluación'

En el contexto del análisis del discurso, una de las primeras descripciones del concepto 'evaluación' se presenta en el trabajo seminal de Labov y Waletzky (1967) referido al análisis del género narrativo. Ambos autores utilizan este término para identificar uno de los cinco componentes estructurales de una secuencia oral[5]. De este modo, la 'evaluación' cumple la función dual de destacar, desde el punto de vista del narrador, el carácter extraordinario de los eventos representados y el valor conversacional del relato[6]. En esta línea, Labov y Fanshel (1977, cit. en Tadros 1994) definen evaluación como "…una expresión de rango superior que incluye acuerdo, disconformidad y tipos de valoraciones que comprenden tanto respuestas cognitivas como evaluativas" (1994:74). En un sentido amplio, Hunston y Thompson (2001) definen este término como "…la expresión de la actitud del hablante o del escritor, o de la posición del hablante, o del punto de vista o de los sentimientos acerca de las entidades o proposiciones [sobre lo cual] él o ella está hablando. Esta actitud puede estar relacionada con la certeza, la obligación, la conveniencia o cualquier otro tipo de valor" (2001:5)[7].

---

[5] Labov y Waletzky (1967) señalan que una narración oral prototípica consta de: a) 'orientación', b) 'acción de complicación de la narración', c) 'evaluación', d) 'resultado' o 'resolución' y e) 'coda' o 'moraleja'.

[6] Labov (1972a:371, cit. en Eggins 2004) complementa esta discusión señalando que "…los recursos evaluativos nos dicen: esto fue aterrador, peligroso, raro, salvaje, loco, entretenido, cómico, maravilloso o, en general, extraño, poco común, casual o digno de contarse". (2004:44)

[7] De aquí en adelante, las citas textuales aquí presentadas son traducciones nuestras.

En otras aproximaciones al término evaluación, Biber et al. (1999) utilizan 'posición del hablante' ('stance') que expresa "...sentimientos personales, actitudes, juicios de valor o evaluaciones" (1999:966). Ochs y Schieffelin (1989) proponen el término 'afecto' "…que incluye sentimientos, estados de ánimo, temperamentos y actitudes asociadas a personas y/o situaciones" (1989:7). Como una particular aproximación al objeto de estudio referido, Martin y White (2005) introducen el término 'valoración' ('appraisal') que comprende "...la presencia subjetiva de los escritores y los hablantes en los textos en tanto adoptan posiciones hacia el contenido del mensaje que presentan como a todos aquellos con quienes se comunican" (2005:1).

3.2 Las categorías formales de evaluación en el nivel textual[8]

Como parte de sus propuestas de un modelo descriptivo que formalice la expresión del componente evaluativo, Hunston y Thompson (2001) señalan la existencia de algunos ítems léxicos cuyo significado semántico y función pragmática principal es la evaluación. Según ambos autores, el conjunto de lexemas evaluativos incluyen "…a) adjetivos: *espléndido(a)*, *terrible*, *sorprendente*, *obvio(a)*, *importante*, *posible*, *falso(a)*, b) adverbios[9]: *felizmente*, *claramente*, *curiosamente*, *posiblemente*, *necesariamente*, c) sustantivos: *éxito*,

[8] Zenteno (1997) señala que "...el término 'textual' implica el reconocimiento del nivel transoracional […] [debiendo] ser interpretado, en primer lugar, como un plano 'interoracional' en que se organiza el contenido proposicional lógico expresado mediante oraciones y cláusulas. En segundo lugar, como 'supraoracional' nivel en que el emisor expresa significados actitudinales que tienen que ver con su propia apreciación, evaluación o modalidad de expresión del contenido factual del mensaje". (1997:170-171)

[9] En el estudio de la expresión de significados evaluativos de esta clase léxica debemos mencionar el trabajo fundacional de Barrenechea (1979) sobre adverbios que funcionan como operadores pragmáticos indicadores de actitud oracional.

fracaso, *tragedia*, *triunfo*, *probabilidad* y d) verbos: *resultar*, *fracasar*, *ganar*, *perder*, *dudar*" (2001:14)[10].

En este sentido, Zenteno (2002-2003) explica que los miembros de las clases léxicas abiertas presentan un mayor número de instancias evaluativas que los miembros de las clases léxicas cerradas[11]. A su vez, Biber et al. (1999) establecen la existencia de marcadores evaluativos (`stance markers`) tanto léxicos como gramaticales. Los primeros incluyen adjetivos, sustantivos y verbos. Los segundos comprenden adverbiales, cláusulas complementarias, verbos modales y semi-modales, entre otros. Con el fin de identificar y categorizar los recursos evaluativos, Hunston y Sinclair (2001:84-91) presentan un conjunto de 6 'esquemas' o 'patrones' ('patterns')[12] léxico-gramaticales interrelacionados a través de los cuales se expresa la evaluación en inglés[13].

1. 'IT' + VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA

		Categoría Evaluativa	Referente Evaluado[14]
'It'	Verbo Relacional	Grupo Adjetivo	Cláusula Finita / no Finita
It	*was*	*certain*	*that he was much to blame.*

[10] Tal como lo indica Tadros (1994), en textos expositivos del ámbito económico, legal y lingüístico la evaluación puede ser expresada mediante el estilo indirecto (i.e., autores citados con referencia a sus teorías) y cuya expresión gramatical corresponde a verbos 'factuales' (e.g., *demostrar*, *probar*, *conocer*) donde el escritor presupone la veracidad de lo expresado. Por el contrario, el autor referido señala que los verbos 'no factuales' (e.g., *sugerir*, *pensar*, *plantear*) establecen una 'imparcialidad' del escritor con lo enunciado que no permite, por lo tanto, la expresión de la evaluación.

[11] Zenteno (2002-2003) da cuenta de la diferencia semántica entre las distintas clases léxicas en el sentido que "…las clases abiertas [i.e., adjetivos, adverbios, sustantivos y verbos] incluyen elementos con significados tanto objetivos como subjetivos. En cambio, los miembros de las clases cerradas [i.e., conjunciones, determinantes, preposiciones, pronombres y verbos auxiliares] (con exclusión, especialmente, de la subclase de verbos auxiliares modales y de un número no menor de otras subclase que también expresan modalidad) tienden a denotar, mayoritariamente, significados objetivos" (2002-2003:312-313). Zenteno (comunicación personal) observa el estatus ambivalente de la clase léxica 'interjección'. Las clasificaciones léxicas (Quirk et al. 1985, Leech et al. 2006) la caracterizan como una clase cerrada, periférica al sistema de la lengua. Sin embargo, puede advertirse su ampliación y/o renovación más o menos permanente de nuevos elementos miembros. Finalmente, sus significados son de carácter subjetivo i.e., expresivos.

[12] Se utilizarán en el curso del presente trabajo, indistintamente, como sinónimos los términos 'patrón' y 'esquema' como traducciones del término inglés 'pattern'.

[13] Para esta revisión bibliográfica hemos mantenido la numeración del original.

[14] De aquí en adelante, traducciones nuestras de los términos en inglés 'Evaluating Context', 'Evaluating Response', 'Hinge' y 'Thing Evaluated'.

2.'THERE' + VERBO RELACIONAL + 'SOMETHING / ANYTHING / NOTHING' + GA + 'ABOUT / IN' + GRUPO NOMINAL / CL *–ING*

'There'	Verbo Relacional	Pivote	Categoría Evaluativa	Pivote	Referente Evaluado
		'something / anything / nothing'	Grupo Adjetivo	'about / in'	Grupo Nominal / Cláusula con *–ing*
There	*'s*	*something*	*rather appealing*	*about*	*being able to spend the evening in a town.*

3. VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA INFINITIVA CON *TO-*

3.1

Referente Evaluado Grupo Nominal	Pivote Verbo Relacional	Categoría Evaluativa Grupo Adjetivo	Restricción de la Evaluación Cláusula Infinitiva con *to-*
Horses	*are*	*pretty*	*to look at.*

3.2

Portador de la Evaluación Grupo Nominal	Pivote Verbo Relacional	Categoría Evaluativa Grupo Adjetivo	Referente Evaluado Cláusula Infinitiva con *to-*
You	*are*	*right*	*to say that.*

3.3

Evaluador Grupo Nominal	Pivote Grupo Verbal	Marca Evaluativa Grupo Adjetivo	Referente Evaluado Cláusula Infinitiva con *to-*
Benjamin	*had been*	*rather overawed*	*to meet one of the Billington family.*

4. VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA CON *THAT-*

4.1

Evaluador Grupo Nominal	Pivote Verbo Relacional	Marca Evaluativa Grupo Adjetivo	Referente Evaluado Cláusula Infinitiva con *that-*
He	*was*	*very angry*	*that she had spoken to people about their private affairs.*

4.2

Portador de la Evaluación Grupo Nominal	Pivote Verbo Relacional	Categoría Evaluativa Grupo Adjetivo	Referente Evaluado Cláusula Infinitiva con *that-*
They	*were*	*lucky*	*that we scored when we did.*

5. Oraciones pseudo-hendidas

5.1

Pivote 'What' + Verbo Relacional	Categoría Evaluativa Grupo Adjetivo	Contexto Evaluativo Frase Preposicional	Pivote Verbo Relacional	Referente Evaluado Cláusula / Grupo Nominal
What's	*very good*	*about this play*	*is*	*that it broadens people's view.*

5.2

Pivote 'What' + Verbo Relacional	Categoría Evaluativa Grupo Adjetivo	Pivote Verbo Relacional	Referente Evaluado Cláusula / Grupo Nominal
What's	*interesting*	*is*	*the tone of the statement.*

5.3

Pivote 'What'	Evaluador Grupo Nominal	Grupo Verbal	Categoría Evaluativa Grupo Adjetivo	Verbo Relacional	Referente Evaluado Cláusula / Grupo Nominal
What	*I*	*find*	*so amazing*	*is*	*that my Dad is a very strict Hindu.*

6. Patrones con sustantivos generales

6.1

Categoría Evaluativa Adjetivo + Sustantivo General	Contexto Evaluativo 'about' + Grupo Nominal	Pivote Verbo Relacional	Referente Evaluado Cláusula / Grupo Nominal
The surprising thing	*about chess*	*is*	*that computers can play it so well.*

6.2

Categoría evaluativa Adjetivo + Sustantivo General	Pivote Verbo Relacional	Referente Evaluado Cláusula / Grupo Nominal
The important point	*Is*	*to involve them in the decision.*

Para los propósitos de nuestro análisis sintáctico ha sido necesario ampliar las categorías del modelo descriptivo de Hunston y Sinclair (2001:84-91). Se ejemplifica, en las siguientes tres secciones[15], cada nuevo esquema propuesto por Prieto y Zenteno (las expresiones evaluativas en subrayado).

[15] Se conserva la numeración original del corpus.

10.1 Artículos de investigación científica

1.1 GRUPO NOMINAL (‘IT’) + VERBO RELACIONAL + GRUPO ADJETIVO
+ CLÁUSULA NOMINAL FINITA / NO FINITA
[GN (‘IT’) + VR + GAdj + Cl] [O Compl]

		Categoría evaluativa	Referente evaluado
‘It’	Verbo Relacional	Grupo Adjetivo	Cláusula Nominal finita / no finita
It	*seems*	*appropriate* […]	*to briefly summarize two major points of interest concerning isochores.* (T. 10.1.1, lín. 67)

1.2 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GN + GV + GN] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal
[…] *the most remarkable ones*	*being*	*the correlations of isochore families* […] (T. 10.1.1, lín. 77)

1.3 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
+ FRASE PREPOSICIONAL
[GN + GV + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal	Frase Preposicional
These findings [ultracentrifugation in Cs2SO4 density gradients in the presence of sequence-specific ligands]	*opened*	*a new inroad*	*in the study of the organization of eukaryotic genomes, superseding DNA reassociation kinetics* […] (T. 10.1.1, lín. 4)

1.4 GRUPO NOMINAL + GRUPO VERBAL + FRASE PREPOSICIONAL
[GN + GV + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Frase Preposicional
Localizing genes in separate isochores	*led*	*to the discovery of an unexpected and strikingly nonrandom distribution of genes* […] (T. 10.1.1, lín. 70)

1.5 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL
+ FRASE PREPOSICIONAL
[FP + GN + GV + FP] [O Simp]

Frase Preposicional	Grupo Nominal	Grupo Verbal	Frase Preposicional
In the last decade,	[…] *several results based on theoretical and experimental approaches*	*led*	*to an improvement of the gene partition criteria* […] (T. 10.1.3, lín. 37)

1.6 GRUPO NOMINAL + GRUPO VERBAL + GRUPO ADVERBIAL
+ CLAÚSULA ADVERBIAL NO FINITA
[GN + GV + GAdv + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal	Grupo Adverbial	Cláusula Adverbial no finita
[…] *they [the findings reported here]*	*go*	*much farther*	*in that they directly identify and map isochores on chromosomes, thus leading to a resolution of >3000 chromosomal bands.* (T. 10.1.1, lín. 31)

1.7 GRUPO NOMINAL + CLÁUSULA ADVERBIAL NO FINITA + GRUPO VERBAL
+ GRUPO NOMINAL + GRUPO ADVERBIAL + CLÁUSULA PREPOSICIONAL
[GN + Cl Adv no fin + GV + GN + GAdv + Cl prep] [O Simp]

Grupo Nominal	Cláusula Adverbial no finita	Grupo Verbal	Grupo Nominal	Grupo Adverbial	Cláusula Preposicional
The present findings,	*while confirming the isochore features previously established,*	*push*	*our knowledge*	*farther*	*by quantifying the size, GC levels, standard deviations, and coordinates of isochores on the human genome map.* (T. 10.1.1, lín. 61)

1.8 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO NOMINAL
[GN + GVpas + GN] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Grupo Nominal
Isochores […]	*have been considered*	*"a fundamental level of genome organization"* […] (T. 10.1.1, lín. 55)

1.9 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA ADVERBIAL
NO FINITA
[GN + GVpas + CL Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal pasivo	Cláusula Adverbial no finita
More studies	*are needed*	*to investigate the underlying mechanism and the nature of this phenomenon.* (T. 10.1.2, lín. 130)

1.10 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL
[GN + GVpas + GAdv] [Cl no fin]

Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial
Gene expression	*is [controlled]*	*more rigidly in germ-line cells than in the surrounding somatic cells.* (T. 10.1.4, lín. 65)
Construcción original: *Gene expression is more rigidly controlled in germ-line cells than in the surrounding somatic cells.*		

1.11 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GN + GVpas + FP] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Frase Preposicional
[…] *the genomes of warm-blooded vertebrates*	*were characterized*	*by a striking long-range compositional heterogeneity* […] (T. 10.1.1, lín. 8)

1.12 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVpas + CL Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal pasivo	Cláusula Nominal no finita
[…] *localized duplication of genomic segments and rearrangement of chromosomal segments*	*have been proposed*	*to be 2 major factors in eukaryotic genome evolution* […] (T. 10.1.2, lín. 8)

1.13 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA NOMINAL NO FINITA + CLÁUSULA ADVERBIAL FINITA
[GN + GVpas + Cl Nom no fin + Cl Adv fin] [O Compl]

Grupo Nominal	Grupo Verbal Pasivo	Cláusula Nominal no finita	Cláusula Adverbial finita
The different computational approaches used to disprove or redefine isochores […]	*were* […] *shown*	*to be inadequate* […]	*even if some of them led to a partial identification of isochores.* (T. 10.1.1, lín. 19)

1.14 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GN + GVrel + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
[…] *"such a selective pressure [the fact that human genome organization and evolution are under selective forces] without apparent correlation with gene expression*	*appeared*	*quite speculative".* (T. 10.1.3, lin. 64)

1.15 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + FRASE PREPOSICIONAL
[GN + GVrel + GAdj + FP] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Frase Preposicional
DNA duplication […]	*is*	*important*	*in adaptive evolution* […] (T. 10.1.2, lín. 1)

1.16 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + CLÁUSULA ADVERBIAL NO FINITA
[GN + GVrel + GAdj + Cl Adv no fin] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Cláusula Adverbial no finita
The apparent rarity of complementation in 'trans'	*may seem*	*surprising*	*given our understanding of the mechanism of retroviral replication.* (T. 10.1.4, lín. 60)

1.17 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO (COMPARATIVO) + CLÁUSULA ADVERBIAL NO FINITA
[GN + GVrel + GAdj (Comp) + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo (Comparativo)	Cláusula Adverbial no finita
[…] *parallel orientation in TAGs*	*appears to be*	*more favored than divergent or convergent orientations* […]	*corroborating Graham's conjecture, at least in the 3 genomes.* (T. 10.1.2, lín. 117)

1.18 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
[GN + GVrel + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
[…] *TAGs*	*are*	*a major component of the genome.* (T. 10.1.2, lín. 41)

1.19 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN + GVrel + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Frase Preposicional
DNA duplication	[…] *is*	*the principle process patterns* […]	*by which the genetic raw material is provided for the origin of evolutionary novelties such as new gene function and expression* […] (T. 10.1.2, lín. 1)

1.20 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + GRUPO ADJETIVO + CLÁUSULA ADVERBIAL/ (CLÁUSULA PREPOSICIONAL)
[GN + GVact + GN + GAdj + Cl Adv / (Cl prep)] [O Compl]

Grupo Nominal	Grupo Verbal Activo	Grupo Nominal	Grupo Adjetivo	Cláusula Adverbial / (Cláusula Preposicional)
This recent activity [the HERV-K family first integrated into the genome of the common ancestor of humans and Old World monkeys at least 30 million years ago]	*makes*	*this family*	*ideal*	*for distinguishing between the alternative mechanisms of proliferation.* (T. 10.1.4, lín. 40)

1.21 GRUPO NOMINAL+ GRUPO VERBAL ACTIVO + GRUPO ADJETIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + GAdj + Cl Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Adjetivo	Cláusula Nominal no finita
The availability of SNP data from many different taxa	[…] *makes* […[it]…]	*feasible*	*to develop a more detailed knowledge of factors that contribute to variation in mutational biases.* (T. 10.1.5, lín. 10)

1.22 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + Cl Nom no fin] [O Comp]

Grupo Nominal	Grupo Verbal activo	Cláusula Nominal no finita
[…] *unequal recombination [of a DNA segment]*	[…] *would cause*	*arrays with oppositely oriented repeats to undergo disastrous duplication-deletion events* […] (T. 10.1.2, lín. 114)

1.23 GRUPO NOMINAL + GRUPO ADVERBIAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[GN + GAdv + GVact + GN] [O Simp]

Grupo Nominal	Grupo Adverbial	Grupo Verbal activo	Grupo Nominal
The paucity of inherited stop codons, and the low dN/dS ratios for all genes (including 'env') within the internal branches of the HERV-K (HML2) phylogeny,	*strongly*	*indicate*	*purifying selection* […] (T. 10.1.4, lín. 49)

1.24 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GAdv + GN + GVrel + GAdj] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
Surprisingly,	*The mechanism by which HERVs have increased in copy number*	*is*	*only poor understood.* (T. 10.1.4, lín. 17)

1.25 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
[GAdv + GN + GVrel + GN] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
Unexpectedly, […]	*regions of increased gene dense*	*are not*	*gene expression.* (T. 10.1.3, lín. 75)

1.26 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GAdv + GN + GV + GN] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal	Grupo Nominal
Interestingly,	*the chromosomes that have greater than expected number of TAG forests*	*tend to have*	*less than expected number of TAGs deserts.* (T. 10.1.2, lín. 84)

1.27 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL + FRASE PREPOSICIONAL
[GAdv + GN + GVpas + GAdv + FP] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal Pasivo	Grupo Adverbial	Frase Preposicional
Until now,	*our understanding of the evolution of interspersed repeats such as HERVs*	*has been influenced*	*heavily*	*by phylogenetic tree shape* […] (T. 10.1.4, lín. 115)

10.2 Artículos periodísticos

1.28 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO (COMPARATIVO)
[GN + GVrel + GAdj (Comp)] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo (Comparativo)
No drugs	*are*	*effective for the other, far more common form of the disease*. (T. 10.2.3, lín. 48)

1.29 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL
+ CLÁUSULA NOMINAL
[GN + GVrel + Cl Nom] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Cláusula Nominal
[…] *our hopes*	*are*	*to do more.* (T. 10.2.1, lín. 61)

1.30 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
+ GRUPO VERBAL ACTIVO + FRASE PREPOSICIONAL
[GN + GVrel + GN + GVact + FP) [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Grupo Verbal activo	Frase Preposicional
[…] *there*	*was*	*interesting stuff*	*going on*	*in these regions [between genes]* […] (T. 10.2.4, lín. 9)

1.31 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + FRASE PREPOSICIONAL
[GN + GVact + FP] [O Simp]

Grupo Nominal	Grupo Verbal activo	Frase Preposicional
No cure or effective treatment	*exists*	*for Alzheimer's* […] (T. 10.2.3, lín. 52)

1.32 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
+ CLÁUSULA NOMINAL NO FINITA
[GN + GVact + GN + Cl Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Cláusula Nominal no finita
Scientists	*expect*	*the rhesus macaque genome sequence*	*to enhance research in neuroscience, behavioral biology, reproductive physiology, endocrinology, and cardiovascular studies.* (T. 10.2.2, lín. 47)

1.33 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
PRONOMINAL + GRUPO ADJETIVO
[GN + GVact + GN (pro) + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal (pronominal)	Grupo Adjetivo
[…] *we*	*'re doing*	*something*	*quite opposite* […] (T. 10.2.2, lín. 79)

1.34 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + GRUPO ADJETIVO + FRASE PREPOSICIONAL + CLÁUSULA ADVERBIAL FINITA / (CLÁUSULA PREPOSICIONAL)
[GN + GVact + GN + GAdj + FP + Cl Adv fin / (Cl prep) [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Grupo Adjetivo	Frase Preposicional	Cláusula Adverbial finita / (Cláusula preposicional)
That divergence	*makes*	*macaques*	*ideal*	*for the evolutionary study of primates,*	*because important features that have been conserved in primates over time can be more easily seen by comparing rhesus monkeys to humans than by comparing chimps to humans.* (T. 10.2.2, lín. 19)

1.35 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + CLÁUSULA INFINITIVA + CLÁUSULA ADVERBIAL / (CLAUSULA PREPOSICIONAL)
[GN + GVact + GN + Cl inf + Cl Adv / (Cl prep)] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Cláusula Infinitiva	Cláusula Adverbial / Cláusula Preposicional
[…] *it [5 percent of the studied DNA sequence]*	*plays*	*an important enough role*	*for evolution to preserve*	*while species have evolved.* (T. 10.2.4, lín. 64)

1.36 GRUPO ADVERBIAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GAdv + GVpas + FP] [O Simp]

Grupo Adverbial	Grupo Verbal pasivo	Frase Preposicional
Little	*is known*	*about the differences between a male and a female genome.* (T. 10.2.1, lín. 56)

1.37 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PRONOMINAL (PREPOSICIONAL) + FRASE PREPOSICIONAL
[GN + GVpas + FPro (prep) + FP] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Frase Pronominal (preposicional)	Frase Preposicional
[…] *the gene*	*is situated*	*somewhere*	*along a particular DNA strand between two easily recognized sequences called markers* […] (T. 10.2.2, lín. 70)

1.38 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + GRUPO ADJETIVO (COMPARATIVO)
[GN + GVrel + GN + GAdj (comp)] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Grupo Adjetivo (comparativo)
[…] *the two subspecies of the macaques*	*are*	*very different from each other on a genetic level,*	*probably much more different than human populations are from each other.* (T. 10.2.2, lín. 58)

1.39 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL + GRUPO NOMINAL
[GN + GVpas + GAdv + GN] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial	Grupo Nominal
[…] *robots and powerful X-ray generators*	*have boosted*	*[lately]*	*the pace of discovery.* (T. 10.2.5, lín. 37)
Construcción original: […] *robots and powerful X-ray generators have lately boosted the pace of discovery.*			

1.40 GRUPO NOMINAL (PRONOMINAL) + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL+ FRASE PREPOSICIONAL
[GN (PRO) + GVrel + GN + FP] [O Simp]

Grupo Nominal (pronominal)	Grupo Verbal relacional	Grupo Nominal	Frase Preposicional
[…] *there*	*is*	*good news*	*for some patients.* (T. 10.2.3, lín. 44)

1.41 CLÁUSULA ADVERBIAL NO FINITA + GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[Cl Adv no fin + GN + GVact + GN] [O Compl]

Cláusula Adverbial no finita	Grupo Nominal	Grupo Verbal activo	Grupo Nominal
[…] *after rigorously pruning the data to keep only the most significant mutations,*	*the researchers*	*identified*	*10 mutations* […] (T. 10.2.1, lín. 29)

10.3 Editoriales

1.42 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN + GVact + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Frase Preposicional
The burgeoning commercial sector that is based on genome information	*poses*	*a challenge*	*to the norms of scientific publication.* (T. 10.3.2, lín. 1)

1.43 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL
[FP + GN + GVpas + GAdv] [O Simp]

Frase Preposicional	Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial
With this shift,	*the burden of genetic disease*	*has changed*	*dramatically.* (T. 10.3.3, lín. 14)

1.44 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL ACTIVO + CLÁUSULA INFINITIVA
[FP + GN + GVact + Cl Adv no fin] [O Compl]

Frase Preposicional	Grupo Nominal	Grupo Verbal activo	Cláusula Adverbial no finita
From a purely scientific standpoint,	*biology research*	*stands*	*to benefit enormously from the availability of the genome sequences of diverse organisms.* (T. 10.3.4, lín. 43)

1.45 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + GRUPO NOMINAL
[GN + GVrel + GAdj + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Grupo Nominal
'Nature'	*is*	*delighted*	*this week to publish the project's analysis.* (T. 10.3.2, lín. 12)

1.46 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + CLÁUSULA INFINITIVA
[GN + GVrel + GAdj + Cl inf] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Cláusula Infinitiva
[…] *it*	*is*	*unethical*	*for a country to push families towards screening.* (T. 10.3.3, lín. 54)

1.47 GRUPO NOMINAL (PRONOMINAL) + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL+ GRUPO ADJETIVO + CLÁUSULA INFINITIVA
[GN (Pro) + GVrel + GN + GAdj + Cl inf] [O Compl]

Grupo Nominal	Grupo Verbal Relacional	Grupo Nominal	Grupo Adjetivo	Cláusula Infinitiva
[…] *there*	*was*	*a splendid storehouse of information*	*already available*	*to provide that resonance.* (T. 10.3.1, lín. 6)

1.48 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + CLÁUSULA NOMINAL NO FINITA
[GN + GVrel + GN + Cl nom fin] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Cláusula nominal finita
It	*is*	*a pity*	*that Professor Modell did not comment on this point* […] (T. 10.3.3, lín. 42)

1.49 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + FRADE PREPOSICIONAL + CLÁUSULA RELATIVA / EXPLICATIVA
[GN + GVrel + GN + FP + Cl rel/expl] [O Simp]

Grupo Nominal	Grupo Verbal Relacional	Grupo Nominal	Frase Preposicional	Cláusula relativa/explicativa
This [the production of the first 'working draft' of the human genome]	*was*	*a landmark*	*in scientific research and a momentous occasion for 'Nature',*	*which was the scientific journal with responsibility for the review and evaluation of the public effort* […] (T. 10.3.5, lín. 4)

1.50 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + CLÁUSULA ADVERBIAL NO FINITA
[GN + GVrel + Cl Adv no fin] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Cláusula Adverbial no finita
The offer	*was*	*for 'Nature' to review the work with the usual rigour bestowed on all of its scientific submissions* […] (T. 10.3.5, lín.39)

1.51 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + CLÁUSULA INFINITIVA
[GN + GVrel + GAdj + CL inf] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Clausula infinitiva
[…] *a claim those authors*	*would be*	*quick*	*to disavow* […] (T. 10.3.1, lín. 8)

1.52 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[GN + GVact + GN] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal
The collated information from these groups	*would provide*	*the most comprehensive analysis of the human genome.* (T. 10.3.5, lín. 15)

1.53 CLÁUSULA NOMINAL NO FINITA + GRUPO VERBAL RELACIONAL
+ GRUPO NOMINAL [Cl nom no fin + GVrel + GN] [O Simp]

Cláusula Nominal no finita	Grupo Verbal relacional	Grupo Nominal
To do the same in eukaryotic cells	*proved*	*a more difficult challenge*. (T. 10.3.1, lín. 41)

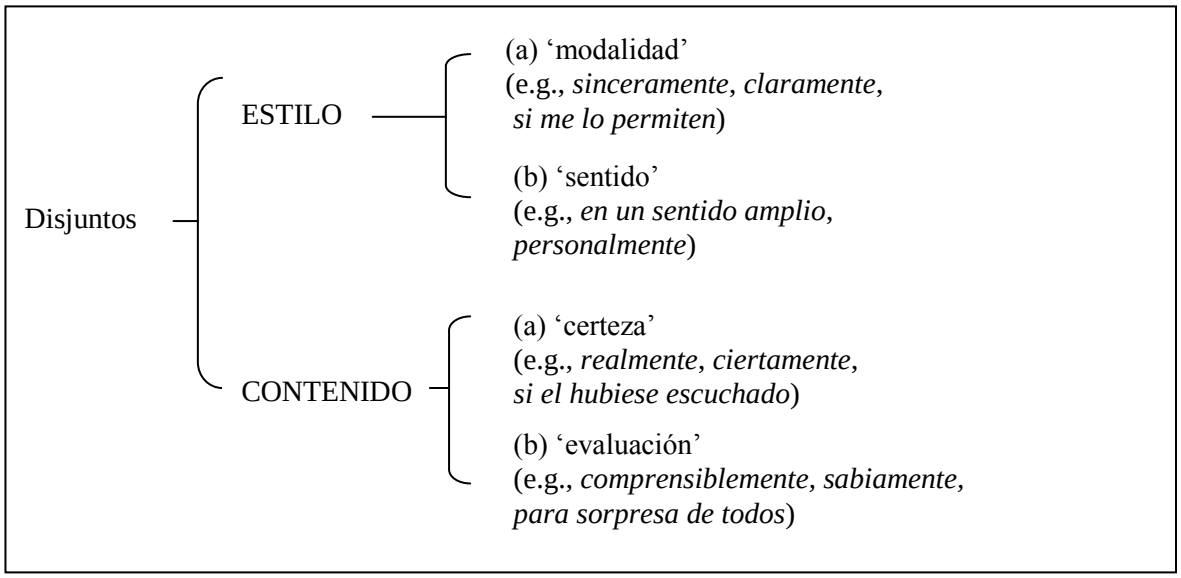
3.3 Los 'disjuntos' como marcadores actitudinales en el discurso

En la expresión de significados evaluativos en el nivel textual, Quirk et al. (1985) definen los disjuntos ('disjuncts') como una subcategoría de adverbiales que "...cumplen un rol superior comparado con los elementos oracionales, por cuanto están sintácticamente más desligados de la estructura oracional y, de alguna forma, son elementos superordinados por cuanto parecen tener un campo de acción que se extiende más allá de los límites de la oración como un todo" (1985:613). Zenteno (1997) señala que éstos "…constituyen una dimensión supraoracional que permite que el escritor o el hablante, exprese, en términos generales, significados de índole subjetiva, relacionados con su actitud frente al contenido proposicional o a su modalidad de expresión" (1997:178).

Por su parte, Quirk et al. (1985) subclasifican a los disjuntos, o 'marcas actitudinales', en dos clases: a) 'estilo' y b) 'contenido' en donde "...los primeros expresan el comentario del hablante sobre el estilo o forma de su expresión mientras que los segundos constituyen observaciones sobre el contenido mismo de la oración y de sus condiciones de veracidad" (1985:615). Como complemento a esta discusión, Zenteno (1997) agrega que "…según el tipo de actitud expresado por el emisor, las marcas de contenido incluyen dos subcategorías: a) 'evaluación' que como su nombre indica, introducen un juicio evaluativo, ya positivo ya negativo ya neutro (e.g., *undestandably*, *to everyone's surprise*, *what is even more remarkable*); y b) 'certeza', que marca, básicamente, una calificación subjetiva del

grado de verdad o de certeza del contenido proposicional de la oración con que se relacionan (e.g., *clearly, obviously, seemingly, it is very likely that*). Por otra parte, en los disjuntos de estilo, Quirk y colaboradores (1985) reconocen dos subcategorías: a) 'modalidad' que explicita el modo de expresión del enunciador (e.g., *truthfully, in short, if I say so without offence*); y b) 'sentido' que delimita el sentido o extensión en que se debe interpretar el contenido proposicional introducido (e.g., *strictly speaking, generally speaking, personally*)" (1997:178). El cuadro 1 reseña las categorías referidas.

Cuadro 1: Disjuntos de estilo y de contenido (Quirk et al. 1985:615, Zenteno 1997:178)



3.4 La teoría de la valoración

La teoría de la valoración ('appraisal theory', Martin 2000a, Macken-Horarik y Martin 2003, Martin y Rose 2003, Martin y White 2005) constituye un modelo descriptivo que da cuenta de los procesos textuales en los cuales los enunciadores expresan y negocian determinadas posiciones personales, intersubjetivas e, incluso, ideológicas. Martin y White (2005) señalan que su modelo, desarrollado a partir de la década de los '80, se relaciona

con "…la presencia subjetiva de los escritores/hablantes en los textos y como éstos adoptan posiciones con respecto del contenido del mensaje y de los interlocutores en la comunicación" (2005:1). A su vez, Martin (2000a) señala que la teoría de la valoración incluye "...el conjunto de recursos [lingüísticos] utilizados para negociar las emociones, juicios y evaluaciones, junto con los recursos para amplificar y comprometerse con las evaluaciones" (2000a:145). Al respecto, Hood y Martin (2005:4) presentan el siguiente ejemplo de enunciado evaluativo (las expresiones valorativas en subrayado).

(1) *"...Éste es el <u>más asombroso</u> de los espectáculos que uno podría llegar a ver aun <u>si a uno no le gustan los blues</u>".*

Por su parte, Martin y White (2005) describen la valoración como un sistema semántico del discurso ya que "…el modelo se concentra en los significados valorativos presentes en un texto" (2005:25). Su teoría clasifica el conjunto de los recursos léxico-sintácticos en tres subsistemas: 'actitud', 'compromiso' y 'gradación'. Dado que el objeto central de nuestro estudio lo constituyen los significados actitudinales en el texto/discurso, a continuación se describe en detalle sólo el primer subcomponente referido.

3.4.1 Actitud

La categoría actitud comprende la expresión del campo semántico de la emoción, el juicio sobre las personas, el comportamiento de éstas y su valoración, especialmente subjetiva, de las cosas del mundo. White (2001) señala que el término actitud incluye los valores que los enunciadores comunican en sus juicios y las respuestas emocionales y afectivas que se asocian con los participantes y los procesos interpretados mediante

evaluaciones positivas o negativas. En cuanto a su clasificación, es posible establecer tres subcategorías: ‘afecto’, ‘apreciación’ y ‘juicio’.

El afecto se vincula con la expresión de la emoción y la emotividad en el lenguaje y comprende los recursos que expresan una respuesta emocional como la felicidad, la tristeza o el temor relacionados directamente con la categoría de disjuntos propuesta por Quirk et al. (1985) y sintetizados en términos de las dicotomías ‘*felicidad/infelicidad*’, ‘*seguridad/inseguridad*’ y ‘*satisfacción/insatisfacción*’ las que se expresan, principalmente, a través de verbos de emoción (e.g., *amar/odiar*), adverbios (e.g., *alegremente*, *tristemente*) y nominalizaciones (e.g., *alegría*/*desesperación*). A continuación, se resume este subsistema (adaptado de Eggins y Slade 1997).

Cuadro 2: Afecto (Eggins y Slade 1997:130)

	AFECTO	
Categoría	Ejemplos positivos	Ejemplos negativos
Felicidad/infelicidad	*Feliz, alegre, jubiloso(a) optimista.*	*Deprimido(a), triste miserable, angustiado(a).*
Seguridad/inseguridad	*Confiado(a), seguro(a), tranquilo(a), sereno(a.).*	*Ansioso(a), preocupado(a), inseguro(a), intranquilo(a).*
Satisfacción/insatisfacción	*Interesado(a), absorto(a), gustar.*	*Cansado(a), aburrido(a), odiar.*

Por su parte, el término apreciación indica la evaluación de las cosas o productos. Martin y White (2005) caracterizan la apreciación como todas aquellas evaluaciones positivas o negativas de objetos, artefactos, procesos y estados en función de principios estéticos. Los valores más comunes en esta subcategoría se relacionan con la forma, la apariencia, la composición, el impacto y la significancia de los artefactos humanos, los objetos físicos, las circunstancias materiales en referencia a la estética y otros sistemas de valor social. Martin (2000a) subclasifica la apreciación en tres dimensiones: a) ‘reacción’,

que describe el impacto y la atención prestada al referente evaluado subdividida en valores de 'impacto' y 'calidad', b) 'composición' relacionada con la percepción de la proporción armónica subdividida en valores de 'balance' y 'complejidad' y c) 'valuación' referida a la apreciación de la importancia social de lo evaluado. White (2001) afirma que la evaluación estética no se limita sólo a los objetos inanimados y estados de las cosas sino que se aplica también a los seres humanos vistos como entidades más que a personas que se comportan individualmente[16]. En la próxima sección, se resume la subcategoría apreciación (adaptada de Martin 2000a:160).

Cuadro 3: Apreciación (Martin 2000a:160)

APRECIACIÓN		
	Positiva	Negativa
REACCIÓN: Impacto	*Cautivador(a), llamativo(a), atractivo(a), agradable, conmovedor(a).*	*Aburrido(a), tedioso(a), ascético(a), pedante, insulso(a).*
REACCIÓN: Calidad	*Hermoso(a), esplendido(a), encantador(a).*	*Feo(a), repulsivo(a), repugnante.*
COMPOSICIÓN: Balance	*Balanceado(a), armonioso(a), simétrico, proporcionado(a).*	*Desbalanceado(a), discordante, desproporcionado(a), asimétrico(a).*
COMPOSICIÓN: Complejidad	*Simple, elegante, detallado(a), preciso(a), intrincado(a).*	*Extravagante, monolítico(a), simplista, impreciso(a).*
VALUACIÓN:	*Profundo(a), innovador(a), original, único(a), exigente.*	*Superficial, insignificante, reaccionario(a), conservador(a).*

White (2001) señala que, en la subcategoría juicio, se utiliza un lenguaje que "…critica o elogia, que condena o aplaude el comportamiento, las acciones, los actos, los dichos, las creencias y las motivaciones de los seres humanos tanto individuales como

[16] Como complemento, Martin y Rose (2003) agregan que "...la apreciación de las cosas incluye nuestras actitudes hacia los espectáculos de televisión, películas, libros, discos compactos; pinturas, esculturas, hogares, edificios públicos, obras de teatro, recitales, desfiles, espectáculos o representaciones de diversa índole; sentimientos acerca de la naturaleza en el caso de vistas panorámicas, amaneceres y crepúsculos, constelaciones, estrellas fugaces y satélites en una noche estrellada". (2003:33)

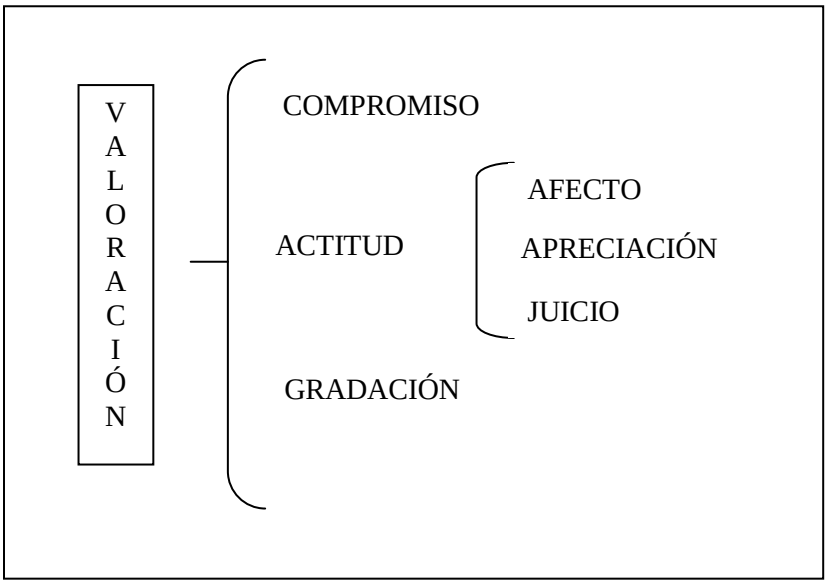
grupales" (2001:1). El juicio se expresa través de: a) la 'estimación social' subdividida en juicios relativos a la 'normalidad', 'capacidad' o 'tenacidad' de una conducta y b) la 'sanción social' constituida por la 'veracidad' e 'integridad moral' en la expresión de la valoración a un nivel social sobre las personas y sus comportamientos, la ética y las normas sociales institucionalizadas. En cuanto a su enunciado lingüístico, éstos corresponden a la clase léxica adjetivos calificativos (e.g., *afortunado*/*desafortunado*), adverbios (e.g., *honestamente*) y verbos (e.g., *engañar*), principalmente. El cuadro 3 ilustra este subcomponente (adaptado de Martin 2000a:156).

Cuadro 4: Juicio (Martin 2000a:156)

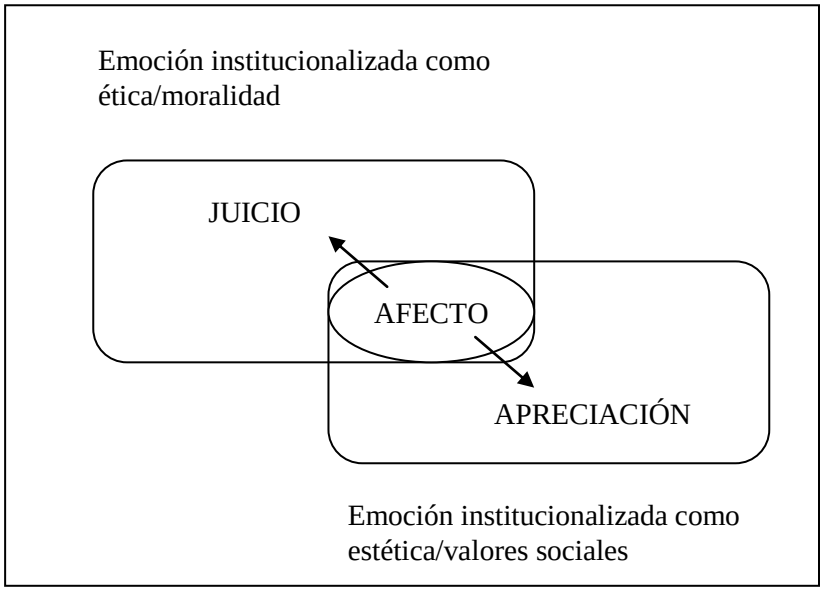
JUICIO		
	ESTIMACIÓN SOCIAL	
	Ejemplos positivos (admiración)	Ejemplos negativos (crítica, sin implicaciones legales)
Normalidad	*Corriente, común, normal, afortunado(a), moderno(a).*	*Excéntrico(a), extraño(a), raro(a), desafortunado(a), anticuado(a).*
Capacidad	*Habilidoso(a), inteligente, intuitivo(a), atlético(a), fuerte.*	*Inhábil, lento(a), tonto(a), torpe, débil.*
Tenacidad	*Heroico(a), valiente, confiable, infatigable, perseverante.*	*Cobarde, apresurado(a), no confiable, distraído(a), perezoso(a).*
	SANCIÓN SOCIAL	
	Ejemplos positivos (alabanza)	Ejemplos negativos (condena, con implicaciones legales)
Veracidad	*Sincero(a), honesto(a), genuino(a), franco(a), directo(a).*	*Deshonesto(a), mentiroso(a), inauténtico(a), manipulador(a).*
Integridad moral	*Moral, bondadoso(a), respetuoso(a) de la ley, sensible, justo(a).*	*Inmoral, malvado(a), corrupto(a), cruel, injusto(a).*

Los cuadros 5 y 6 establecen la relación entre la actitud y los restantes componentes de la teoría de la valoración.

Cuadro 5: Teoría de la valoración (adaptado de Hood y Martin 2005:7)



Cuadro 6: Subsistemas semánticos de afecto, apreciación y juicio (Martin 2000a:147)



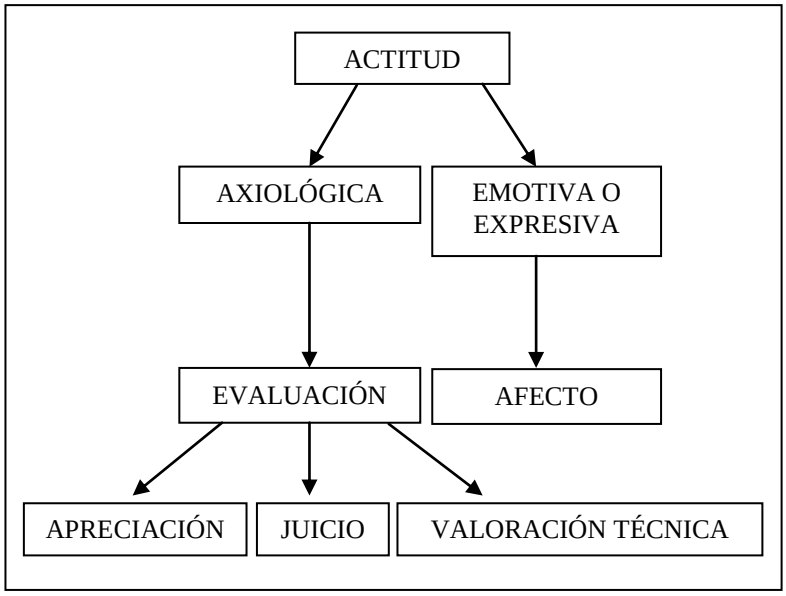
3.4.2 Propuesta de Soto y Zenteno (2010)

En relación con la reformulación parcial de algunos componentes de la teoría de la valoración, Soto y Zenteno (2010) señalan que la subcategoría apreciación no da cuenta en forma exhaustiva de un subconjunto de referentes relacionados con "…el rendimiento físico, intelectual, científico o artístico, ya personal o grupal" (2010:341). Para tal efecto,

los autores referidos, en forma conjunta, proponen el término ‘valoración técnica’ que comprende “…la evaluación de los ‘artefactos’, vale decir, productos o procesos que no solo se originan en el dominio de las artes, o en campos de acción relacionados, sino también en el dominio científico, el tecnológico, el económico o el político[17]” (2010:342).

Soto y Zenteno (2010), en su reformulación conceptual, vinculan el afecto con la expresión del dominio ‘emocional’ y ‘expresivo’ del enunciador. Del mismo modo, asignan a los restantes subcomponentes un valor ‘axiológico’ relacionado con códigos estéticos y normas sociales para la apreciación; la calificación de acciones, productos o artefactos humanos para la valoración técnica y los principios éticos o morales para la expresión de juicio. El siguiente cuadro resume las modificaciones referidas.

Cuadro 7: Reformulación de Soto y Zenteno (2010)



3.5 El discurso expositivo científico

En las últimas décadas, los estudios referidos a la comunicación científica caracterizan al discurso expositivo científico como objetivo, especializado, que informa un tipo especial de conocimiento legitimado por la comunidad discursiva de origen (Harvey

[17] Con especial énfasis en este punto, Soto y Zenteno (2010) agregan que en el ámbito de la política “…las acciones emprendidas o los resultados logrados son sometidos a evaluación regular, sobre la base de parámetros como eficiencia versus ineficiencia, satisfacción versus insatisfacción, buena versus mala calidad de funcionamiento de los productos o de los servicios públicos, entre otros”. (2010:341)

2005) y expresado a través de un lenguaje neutral e impersonal con total ausencia de emociones, sentimientos y apreciaciones (García Negroni 2008). Tal impersonalidad, según Calsamiglia y Tusón (1999) ha sido generada a partir de "...imponer condiciones racionales a la visión de la realidad (meta del conocimiento científico) con el fin de "…dar cuenta de forma organizada y exhaustiva de las características del mundo natural y social" (1999:281).

Así, el enunciador se convierte en un sujeto que conoce el tema, mientras que el receptor es el individuo que será informado sobre ese contenido a través del uso de "...un lenguaje objetivo, con léxico preciso y objetivo" (Cassany et al. 1994:339) estableciendo entre ambos una relación 'asimétrica' donde el primero posee mayor información que el segundo acerca del tópico comunicado. Propios de este discurso especializado son el 'artículo de investigación científica', el 'artículo de divulgación científica' y el 'editorial de prensa' cuyas características serán reseñadas a continuación.

3.5.1 Artículo de investigación científica

Diversos autores consideran al artículo de investigación científica como el texto prototípico del discurso científico (Halliday y Martin 1993, Swales 1990[18], 2004). En general, se caracteriza a este discurso especializado como 'persuasivo' (Hyland 2008)[19] en tanto el enunciador trata de convencer a la comunidad científica de la validez de los resultados de su trabajo de investigación (Harvey 2005). Swales (1990) identifica en su modelo 'IMRD' ('Introducción', 'Método', 'Resultados' y 'Discusión') su organización

[18] Swales (1990) califica al artículo de investigación científica como "...el producto principal de la industria productora del conocimiento". (1990:125)

[19] Hyland (2008) señala que este tipo discursivo comunica "...verdades que emergen de nuestro acceso directo al mundo exterior. El texto es simplemente el canal mediante el cual los científicos informan sobre los hechos observables". (2008:2)

retórica señalando que la 'Introducción' constituye la sección donde el enunciador intenta persuadir a la comunidad académica acerca del valor de su investigación destacando, por cierto, que su trabajo pretende establecer un nuevo nicho de investigación que llene un espacio vacío[20].

Con el fin de ilustrar lo anterior, Swales (2004:227) presenta su modelo de tres movimientos retóricos, desarrollado en la década de los noventa, denominado 'modelo CARS' ('Create-A-Research-Space model') para la sección Introducción de los artículos de investigación en inglés, el cual es comparado con la variación terminológica propuesta por Lewin, Fine y Young (2001, cit. en Swales 2004:235).

Cuadro 8: Modelo de Swales (1990) cf. Lewin, Fine y Young (2001)

	Swales (1990)	Lewin, Fine y Young (2001)
M1	Establecer un territorio.	Reclamar la pertinencia del campo de estudio.
	↓	↓
M2	Establecer un nicho.	Establecer la brecha que la investigación debe completar.
	↓	↓
M3	Ocupar el espacio.	Examinar los resultados obtenidos por el autor.

Tal como establece el carácter persuasivo de la sección Introducción, Ferrari (1999) destaca el mismo valor para la sección 'Conclusiones' de esta clase textual donde se pretende "…resaltar los descubrimientos realizados y situarlos en el contexto de un cuerpo de conocimientos más amplio para poder relacionarlos con el trabajo de otros especialistas" (1999:10).

[20] En relación con este punto, Soto y Zenteno (2001-2003) sostienen que esta sección "…se caracteriza por presentar una compleja articulación interna que obedece a criterios semántico-pragmáticos como la entrega sumaria de contenidos globales del texto, y retóricos, como la inserción eficaz de la investigación en el campo de estudios pertinente". (2001-2003:35)

3.5.2 Artículo de divulgación científica

En su forma canónica, este tipo de comunicación científica presenta información significativa sobre teorías o descubrimientos científicos relatados en forma cronológica (Myers 1994) y construidos a través de un vocabulario preciso y disciplinar con el fin de dar mayor objetividad al texto, cuyos lectores constituyen el público no especialista. En esta línea, Harvey (2005) caracteriza al artículo de divulgación científica como "…un discurso sobre otro discurso" donde el escritor "…recontextualiza su aprensión del discurso fuente y su valoración de los hechos informados" (2005:98). El siguiente cuadro de Harvey (2005) resume y contrasta las características del 'artículo especializado' con el 'artículo de divulgación científica'.

Cuadro 9: Artículo especializado cf. artículo de divulgación científica (Harvey 2005:97-98)

Artículo especializado	Artículo de divulgación científica
El texto como tal representa un acto de investigación realizado por un sujeto observante.	*El texto como tal representa un acto de comunicación realizado por un sujeto participante.*
En el texto se construye un argumento; el argumento es atemporal.	*En el texto se construye una historia; la historia es cronológica.*
Las proposiciones que sustentan el argumento son representaciones que se construyen en mundos posibles.	*Los hallazgos científicos son presentados como hechos pertenecientes al mundo real.*
La palabra del 'otro' es un recurso para construir el argumento y sustentar la argumentación.	*La alteridad es un recurso que le otorga credibilidad a la historia y validez a la información.*
En las proposiciones se tematiza y valoriza el objeto de estudio y/o la actividad investigativa.	*En el texto se tematiza y valoriza la fuente informativa, animada o inanimada y/o la relación ciencia/sociedad.*
El valor de verdad de las proposiciones es siempre provisional, abierto a discusión y sujeto a cuestionamiento.	*Los hechos científicos no se discuten y su verosimilitud no está sujeta a cuestionamiento.*
Los resultados del acto de investigación se evalúan, de preferencia, mediante expresiones de la modalidad epistémica.	*El objeto de la comunicación se evalúa, de preferencia, por medio de la modalidad deóntica.*
Se recurre con frecuencia a nominalizaciones que crean entidades.	*Se utilizan, de preferencia, elementos verbales que crean aconteceres.*
Se privilegian las cláusulas sin agente explícito, encontrándose su mayor densidad en las	*Se observa presencia de marcadores de persona en posición de sujeto gramatical en*

secciones 'método' y 'resultados'.	*todas las secciones del texto, pero no necesariamente en todas las cláusulas.*
Se acuñan nuevos términos y/o se utilizan tecnolectos de significado unívoco en la disciplina.	*Cuando se utilizan términos especializados, estos se explican mediante metáforas, analogías o frases explicativas.*
Se utilizan recursos icónicos y gráficos para construir o sustentar el argumento.	*Cuando se incluyen recursos icónicos y gráficos se usan con fines explicativos.*

3.5.3 Editorial de prensa

Como texto argumentativo, el editorial de prensa es un artículo de opinión frecuentemente escrito por un especialista y referente a una problemática determinada. Bolivar (1997) complementa esta discusión señalando que "…el texto editorial, por convención, tiene la función social de evaluar los acontecimientos y los estados de cosas del mundo y, por lo tanto, constituye uno de los géneros predilectos para la expresión de prácticas discursivas que atañen a la transmisión de valores y opiniones" (1997:9). Con el fin de establecer la expresión del componente evaluativo en las enunciación de la opinión personal del editorialista, Zenteno (2002-2003) puntualiza que en esta clase textual "…el especialista examina tal estado de cosas para proceder a expresar una serie de opiniones que contienen juicios apreciativos tanto de la problemática en general como de aspectos específicos de ella. Mediante tales opiniones, el enunciador pone de manifiesto su particular evaluación del problema, sustentándola ya sea sobre la base de su conocimiento de experto, o como parece ser el caso, en evidencia empírica derivada de su propia investigación" (2002-2003:307).

4. HIPÓTESIS

4.1 Hipótesis central

El componente evaluativo, que en forma característica se hace presente en las clases textuales propias del discurso argumentativo y narrativo, también se encuentra presente en tipos textuales[21] científicos, los que son representativos del discurso expositivo escrito en inglés.

4.2 Hipótesis nula

Los textos científicos escritos en inglés, instancias de discurso expositivo, se caracterizan por presentar una estructura expositiva propia de la investigación científico-empírica, la cual no favorece la ocurrencia del componente evaluativo-valorativo del discurso.

5. PREGUNTAS DE INVESTIGACIÓN

5.1 ¿Qué patrones sintácticos evaluativos presentes en textos expositivos científicos escritos en inglés (según las propuestas teórico-descriptivas de Hunston y Sinclair 2001), demuestran la mayor frecuencia de ocurrencia?

5.2 ¿Qué significados valorativos presentes en textos expositivos científicos escritos en inglés (según las propuestas teórico-descriptivas de Martin y White 2005), demuestran la mayor frecuencia de ocurrencia?

[21] En este punto seguimos a Ciapuscio (1994).

6. METODOLOGÍA

6.1 Descripción del corpus

El presente estudio comprendió el análisis de un corpus consistente en 15 textos expositivos científicos escritos en inglés y correspondientes al área disciplinar de las ciencias médicas y biológicas. Para tal efecto, se seleccionaron 5 artículos publicados en revistas científicas, 5 artículos periodísticos de investigación científica publicados en revistas destinadas al público en general y 5 editoriales publicados en números extraordinarios de revistas de investigación científica. El número de palabras en el corpus correponde a un total de 18,078 desglosado como sigue: a) 8,513 palabras en los artículos de investigación científica, b) 4,829 en los artículos periodísticos y c) 4,736 en los editoriales, con un promedio de palabras por tipo de texto de 1,703, 948 y 966, respectivamente.

Los primeros, fueron seleccionados de las publicaciones 'Gene' ('Expression patterns and gene distribution in the human genome'), 'Genetics' ('Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition'), 'Genome Research' ('An isochore map of human chromosomes'), 'Molecular Biology and Evolution' ('A roundmap of tandemly arrayed genes in the genome of human, mouse, and rat') y 'Proceedings of the National Academy of Sciences of the United States of America' ('Long-term reinfection of the human genome by endogenous retroviruses').

Los segundos, fueron presentados en las revistas 'Discover' ('Researchers focus on differences between groups to find bad DNA', 'National Geographic News' ('Macaque genome deciphered; may herald medical breakthroughs', 'Science News' ('First complete

cancer genome sequenced'), 'Scientific American' ('The 1 percent genome solution') y 'Time Magazine' ('The next frontier: proteomics').

Los terceros, fueron publicados en las revistas 'Bulletin of the World Health Organization' ('Our human genome -how can it serve us well?'), 'Current Science', ('Post-genome blues'), 'Human Genome Collection' ('The 'finished' landscape'), 'Nature' ('Human genomes, public and private') y 'Science' ('DNA: one teacher's reflection').

Los textos fueron seleccionados entre los años 2000 y 2008, todos los cuales son versiones electrónicas recolectadas de los correspondientes sitios web de las publicaciones referidas. El corpus recopilado presenta la discusión del denominado 'Proyecto Genoma Humano ('PGH'), consorcio público internacional de cooperación científica bajo la dirección, desde el año 1990, del biólogo estadounidense James Watson. El propósito central del proyecto, de acuerdo con Yankovic (2007), es "…determinar la secuencia completa de los genes de la especie humana, para conocer su localización y función. Adicionalmente, el PGH incluye el estudio de algunos organismos usados en biología para experimentación de laboratorio, como la mosca 'Drosophila' y la levadura de cerveza" (2007:138).

En una aproximación conjunta, el PGH fue desarrollado por el Departamento de Energía de los Estados Unidos y el Instituto Nacional de Salud de los Estados Unidos de Norteamérica, junto con científicos de Alemania, China, Francia, India, Reino Unido y Japón. Casi en forma paralela al inicio del PGH, el científico estadounidense Craig Venter fundó, en 1998, la empresa privada 'Celera Genomics', para llevar a cabo su propio Proyecto Genoma Humano en 1999 con propósitos comerciales, al margen del PGH. Finalmente, el PGH fue completado en forma parcial en 2001 y en forma total, en 2003.

6.2 Criterios para la recolección y selección final del corpus

La elección del área disciplinar y del corpus de estudio se basó en tres criterios operacionales:

a) el tópico discursivo, polémico y transversal, permite la identificación, mediante las perspectivas teóricas-analíticas referidas, de numerosas instancias evaluativas y valorativas;

b) los textos temáticos afines permiten identificar, categorizar, determinar y comparar los resultados obtenidos; y

c) el corpus seleccionado proporciona un nicho de investigación en el área de estudio propuesta con el fin de establecer un modelo descriptivo preliminar del componente evaluativo en textos expositivos científicos escritos en inglés.

6.3 Procedimientos de análisis

6.3.1 Aplicando las propuestas teórico-descriptivas de Hunston y Sinclair (2001) se identificó[22] y categorizó, mediante un análisis sintáctico, la ocurrencia de los patrones sintácticos evaluativos presentes en artículos periodísticos y textos editoriales examinados. En el caso de los artículos de investigación, se analizó sólo la sección 'Introducción' y 'Discusión de los Resultados', siguiendo las indicaciones de Ferrari (1999) y Swales (1990)[23] al respecto.

[22] El análisis sintáctico de las expresiones evaluativas ha sido realizado estableciendo, como límite descriptivo, la estructura sintáctica de la cláusula (u oración), sea ésta simple o compleja. Por tanto, el análisis demuestra que una expresión (o expresiones evaluativas) puede(n) ser formalizada(s) mediante, en su calidad de constituyente oracional, un grupo (o frase) o cláusula. En relación con este último término, se seguirá la terminología de Halliday y Mathiessen (2004), quienes emplean el concepto 'cláusula' en lugar de 'oración'.

[23] Se remite a la sección 3.5 del presente documento, donde se fundamenta tal elección.

6.3.2 Aplicando las propuestas teórico-descriptivas de Martin y White (2005), se identificó y categorizó, mediante un análisis semántico-pragmático, la ocurrencia de los significados valorativos presentes en los textos examinados.

6.3.3 Se determinó, mediante un procesamiento estadístico, la frecuencia de ocurrencia de tanto los patrones sintácticos evaluativos como de los significados valorativos presentes en los textos referidos.

6.3.4 Teniendo como referente tanto los modelos analíticos propuestos al inicio de este estudio como de sus reformulaciones parciales se compararon las modalidades de funcionamiento de los patrones evaluativos y los significados valorativos en el corpus bajo estudio.

7. PRESENTACIÓN Y DISCUSIÓN DE LOS RESULTADOS

En esta sección, se presentan, ilustran y discuten los resultados obtenidos del análisis, tanto de los patrones sintácticos evaluativos como de los significados valorativos identificados en el corpus bajo estudio.

Tabla 1: Patrones evaluativos en ARCA[24]

	N° texto					
Esquema	10.1.1	10.1.2	10.1.3	10.1.4	10.1.5	TOTAL
1.1	1	0	1	0	1	3
1.2	1	1	4	0	3	9
1.3	1	0	0	0	0	1
1.4	1	1	0	0	0	2
1.5	0	0	1	0	0	1
1.6	1	0	0	0	0	1
1.7	1	0	0	0	0	1
1.8	1	0	0	0	0	1
1.9	0	1	0	0	0	1
1.10	0	0	0	1	0	1
1.11	1	0	0	0	0	1
1.12	0	1	0	0	0	1
1.13	1	0	0	0	0	1
1.14	0	0	1	1	1	3
1.15	0	2	0	0	1	3
1.16	0	0	0	1	0	1
1.17	0	1	0	0	0	1
1.18	0	4	2	0	2	8
1.19	0	2	0	0	0	2
1.20	0	0	0	1	0	1
1.21	0	0	0	0	2	2
1.22	0	1	0	0	0	1
1.23	0	0	0	2	0	2
1.24	0	0	0	1	0	1
1.25	0	0	1	0	0	1
1.26	0	1	1	0	0	2
1.27	0	0	0	1	0	1
SUBTOTAL	9	15	11	8	10	53
TOTAL						

[24] Abreviación del término ‘Artículos publicados en Revistas Científicas’.

Tabla 2: Patrones evaluativos en APIC[25]

	N° texto					
Esquema	10.2.1	10.2.2	10.2.3	10.2.4	10.2.5	TOTAL
1.1	2	1	0	0	0	3
1.2	1	3	2	1	2	9
1.3	0	0	0	0	1	1
1.11	0	0	1	0	0	1
1.14	3	0	1	0	0	4
1.15	0	0	1	1	0	2
1.16	0	0	1	0	0	1
1.17	0	0	1	0	0	1
1.18	1	2	3	1	1	8
1.20	0	1	0	0	0	1
1.21	1	0	0	0	0	1
1.28	0	0	2	0	1	3
1.29	1	0	0	0	0	1
1.30	0	0	0	1	0	1
1.31	0	0	1	0	0	1
1.32	0	1	0	0	0	1
1.33	0	1	0	0	0	1
1.34	0	1	0	0	0	1
1.35	0	0	0	1	0	1
1.36	1	0	0	0	0	1
1.37	0	1	0	0	0	1
1.38	0	1	0	0	0	1
1.39	0	0	0	0	1	1
1.40	0	0	1	0	0	1
1.41	1	0	0	0	0	1
SUBTOTAL	9	12	14	5	6	46
TOTAL						

[25] Abreviación del término ‘Artículos Periodísticos de Investigación Científica publicados en Revistas destinadas al público en general’.

Tabla 3: Patrones evaluativos en ERI[26]

	N° texto					
Esquema	10.3.1	10.3.2	10.3.3	10.3.4	10.3.5	TOTAL
1.1	0	1	2	1	2	6
1.2	3	0	0	0	0	3
1.4	1	0	0	0	0	1
1.8	0	0	0	0	1	1
1.11	0	0	0	0	1	1
1.14	0	2	0	0	0	2
1.18	0	2	1	2	0	5
1.19	0	1	0	0	0	1
1.23	0	1	0	0	1	2
1.31	0	0	1	1	0	2
1.42	0	1	0	0	0	1
1.43	0	0	1	0	0	1
1.44	0	0	0	1	0	1
1.45	0	2	0	0	0	2
1.46	0	0	2	0	0	2
1.47	1	0	0	0	0	1
1.48	0	0	1	1	0	2
1.49	0	0	0	0	1	1
1.50	0	0	0	0	1	1
1.51	1	0	0	0	0	1
1.52	0	0	0	0	1	1
1.53	1	0	0	0	1	2
SUBTOTAL	7	10	8	6	9	40
TOTAL						

[26] Abreviación del término ‘Editoriales publicados en Números Extraordinarios de Revistas de Investigación Científica’.

Tabla 4: Categorías valorativas en ARCA

Clase	Subclase	N° texto					TOTAL
		10.1.1	10.1.2	10.1.3	10.1.4	10.1.5	
Evaluador	EVext	6	4	2	0	1	13
	EVint	3	11	9	9	10	42
Tipo de valoración	AFE	0	1	0	0	0	1
	APR	0	0	0	0	0	0
	JUI	0	0	0	0	0	0
	VT	9	14	11	9	11	54
Tipo de evaluación	(+)	7	3	6	1	2	19
	(-)	1	6	4	1	4	16
	(N)	1	9	1	7	5	23
SUBTOTAL		27	48	33	27	33	168
TOTAL							

Tabla 5: Categorías valorativas en APIC

Clase	Subclase	N° texto					TOTAL
		10.2.1	10.2.2	10.2.3	10.2.4	10.2.5	
Evaluador	EVext	6	5	1	2	0	14
	EVint	5	7	11	6	6	35
Tipo de valoración	AFE	2	0	6	0	2	10
	APR	0	0	0	0	0	0
	JUI	0	0	0	0	0	0
	VT	9	12	6	8	4	39
Tipo de evaluación	(+)	6	6	3	2	3	20
	(-)	2	0	8	1	1	12
	(N)	3	6	1	5	2	17
SUBTOTAL		33	36	36	24	18	147
TOTAL							

Tabla 6: Categorías valorativas en ENER

Clase	Subclase	N° texto					TOTAL
		10.3.1	10.3.2	10.3.3	10.3.4	10.3.5	
Evaluador	EVext	0	1	0	0	0	1
	EVint	7	9	8	6	8	38
Tipo de valoración	AFE	1	1	4	1	1	8
	APR	0	0	0	0	0	0
	JUI	0	2	2	0	1	5
	VT	6	7	2	5	6	26
Tipo de evaluación	(+)	2	4	3	3	8	20
	(-)	4	1	4	3	0	12
	(N)	1	5	1	0	1	8
SUBTOTAL		21	30	24	18	24	118
TOTAL							

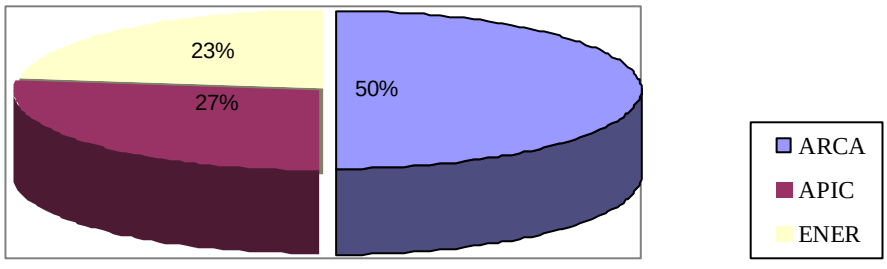
Tabla 7: Construcciones sintácticas funcionales en el nivel textual

Tipo de oración	Tipo de texto			TOTAL
	ARCA	ENER	APIC	
Oración simple	40	28	37	105
Oración compleja	12	12	12	36
Cláusula	1	0	0	1
SUBTOTAL	53	40	49	142
TOTAL				

7.1 Resultados

El análisis de los datos dio como resultado un número total de 53 patrones sintácticos que corresponden a 139 instancias evaluativas identificadas en el corpus examinado. Consecuentemente, 127 de éstas se adscriben a la propuesta de Prieto y Zenteno y sólo 12 al modelo de Hunston y Sinclair (2001). De acuerdo con este último modelo referido, sólo el esquema 1.1 [GN (‘IT) + VR + GAdj + CL] se ha identificado en el corpus: 12 instancias, con ocurrencias en los artículos de investigación científica, artículos periodísticos y editoriales como sigue: 3 (25%), 3 (25%) y 6 (50%), respectivamente. En un claro contraste, según la propuesta de Prieto y Zenteno, de las 127 intancias y 52 esquemas restantes, el orden de éstos últimos decrece como sigue: 26 (50%) de éstos fueron encontrados en los artículos de investigación científica, 14 (27%) fueron identificados en los artículos periodísticos y 12 (23%) fueron determinados en los editoriales. Este último análisis se presenta en el gráfico siguiente.

Gráfico 1: Frecuencia de ocurrencia de patrones evaluativos



En los artículos de investigación científica, el esquema 1.2 [GN + GV + GN] registra el mayor número de ocurrencias (9) seguido del patrón 1.18 [GN + GVrel + GN] con 8 instancias. El mismo fenómeno se aprecia en los artículos periodísticos, donde el esquema 1.2 concentra la mayor frecuencia de ocurrencia (9 casos) superando levemente al esquema 1.18, con 8 instancias. Por otra parte, en los editoriales, se observa un leve predominio de los patrones 1.1 y 1.18 (6 y 5 ocurrencias, respectivamente). A continuación, se ejemplifican los dos esquemas con mayor frecuencia de ocurrencia en cada tipo de texto (las expresiones evaluativas en subrayado).

(2) […] *genes located in the GC-rich isochores required highly efficient translation* […][27]

(3) […] *TAGs are a major component of the genome*.[28]

(4) *Tiny slice of genome reveals bustling activity in the gaps between genes*.[29]

(5) *Some interesting findings are already coming from a study of variations within the macaque genome*.[30]

(6) […] *it is encouraging that in the UK there were no problems about confidentiality* […][31]

(7) *The newspapers were full of headlines screaming of the benefits of the project to humanity*.[32]

En relación con las categorías valorativas, el análisis permitió la identificación de 143 instancias cuantificadas en un orden decreciente: 55 (39%) en los artículos de investigación científica, 49 (35%) en los artículos periodísticos y 39 (26%) en los editoriales. En este

[27] Patrón 1.2, T.10.1.3, lín. 18.

[28] Patrón 1.18, T.10.1.2, lín. 41.

[29] Patrón 1.2, T.10.2.3, lín. 1.

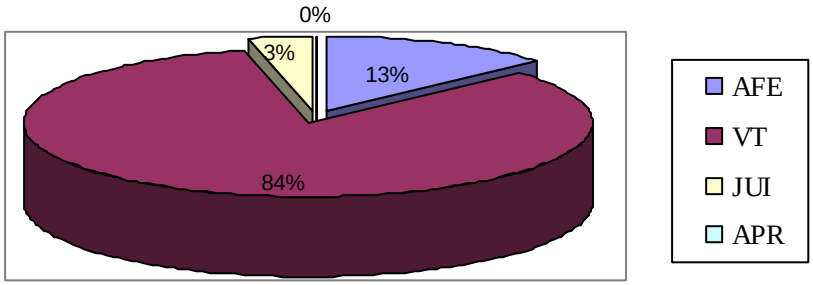
[30] Patrón 1.18, T.10.2.2, lín. 51.

[31] Patrón 1.1, T.10.3.3, lín. 44.

[32] Patrón 1.18, T.10.3.4, lín. 3.

análisis se observa el predominio de la subcategoría ‘valoración técnica’, seguidas de las subcategorías ‘afecto’ y ‘juicio’. En total, en 119 de los 143 casos, los enunciados se asocian a la primera categoría mencionada (84%); en 19 de ellos, se expresa afecto (13%), mientras que en sólo 5, que corresponde a un 3%, se expresa juicio. El siguiente gráfico resume este último análisis cuantitativo.

Gráfico 2: Frecuencia de ocurrencia de categorías valorativas



A su vez, de un total de 119 enunciados clasificados como valoración técnica, 54 (45%) se ubican en los artículos de investigación científica, 39 (33%) en los artículos periodísticos y 26 (22%) en los editoriales. Asimismo, este último tipo textual concentra las únicas 5 instancias de juicio en el corpus. De igual forma, en los artículos periodísticos y los editoriales se observa una proporción casi idéntica de instancias de afecto: 10 (53%) versus 8 (42%), respectivamente. Los siguientes gráficos resumen lo descrito.

Gráfico 3: Instancias de valoración técnica por tipo de texto

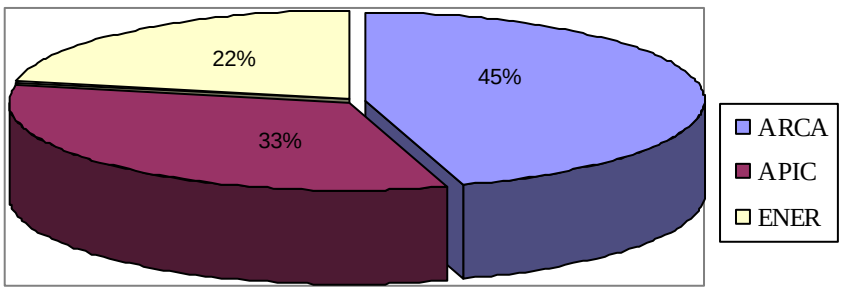


Gráfico 4: Instancias de afecto por tipo de texto

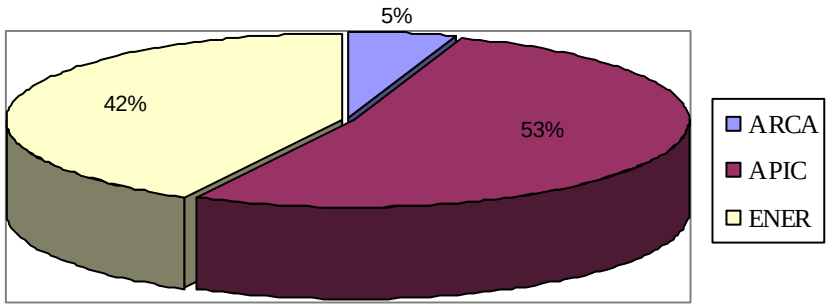


Gráfico 5: Instancias de juicio por tipo de texto

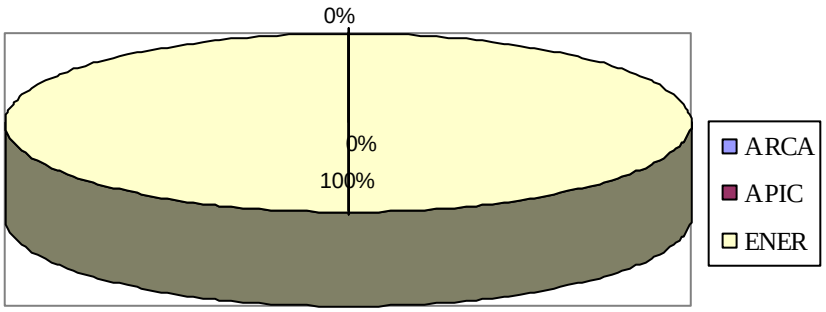
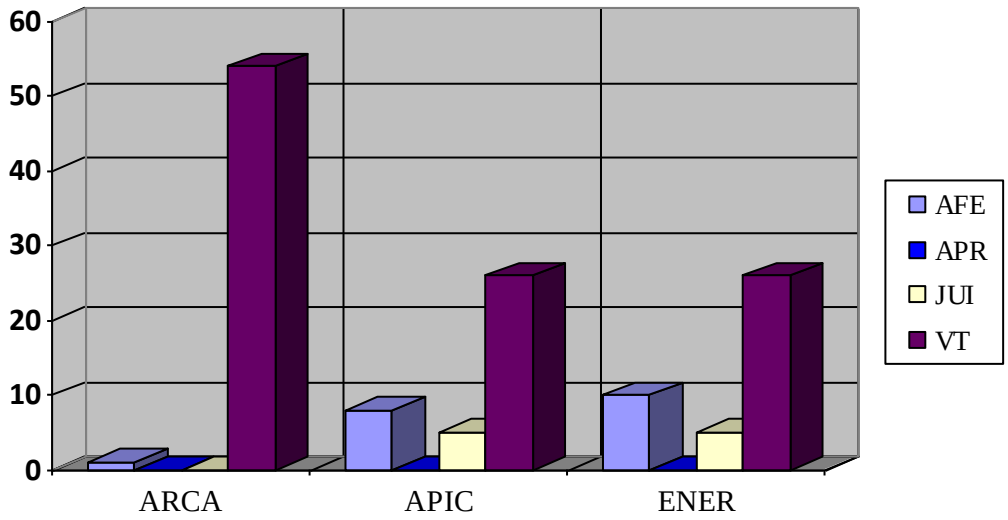


Gráfico 6: Resumen esquemas valorativos por tipo de texto



A modo de ejemplo, en lo que viene, se ilustran cada una de las subcategorías actitudinales referidas (las expresiones evaluativas están subrayadas).

(8) [AFE (+)] *It's fun to speculate* […][33]

(9) [AFE (-)] *No cure or effective treatment exists for Alzheimer's* […][34]

(10) [AFE (N)] *This issue of 'Science' appears amid a swirl of contemporary reminiscence about DNA* […][35]

(11) [JUI (+)] *In keeping with the spirit of the genomics community work ethic, 'Nature' has made each of these papers freely available on the Internet*.[36]

(12) [JUI (-)] […] *it is unethical for a country to push families towards screening* […][37]

(13) [JUI (N)] *The burgeoning commercial sector that is based on genome information poses a challenge to the norms of scientific publication*.[38]

(14) [VT (+)] *Localizing genes in separate isochores led to the discovery of an unexpected and strikingly nonrandom distribution of genes* […][39]

(15) [VT (-)] […] *locating on different strands is detrimental to the stability of the array* […][40]

(16) [VT (N)] […] *TAGs are a major component of the genome*.[41]

Con respecto al evaluador, se observa un claro predominio de la evaluación interna por sobre la externa, con 115 (88%) y 28 (12%) ocurrencias, respectivamente. En el análisis por tipo de texto, los artículos de investigación científica registran la mayor cantidad de evaluación interna: (42 casos, 37%), seguidos de los editoriales y los artículos periodísticos: 38, con un 33% y 35 instancias, con un 30%, respectivamente. En relación con el tipo de evaluación externa identificada en el corpus, las ocurrencias se reparten en orden

[33] T.10.2.1, lín. 43.

[34] T.10.2.3, lín. 52.

[35] T.10.3.1, lín. 1.

[36] T.10.3.5, lín. 20.

[37] T.10.3.3, lín. 54.

[38] T.10.1.2, lín. 1.

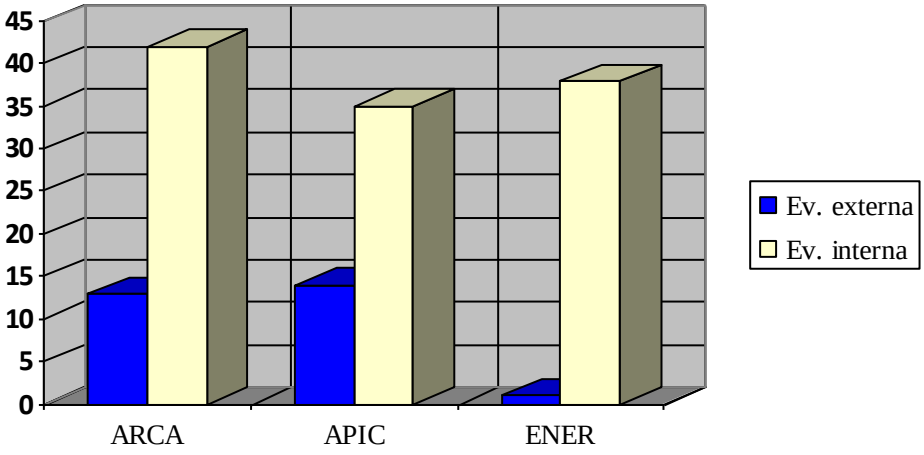
[39] T.10.1.1, lín. 70.

[40] T.10.1.2, lín. 31.

[41] T.10.1.2, lín. 41.

decreciente como sigue: en los artículos periodísticos 14 (50%), en los artículos de investigación científica 13 (46%) y sólo 1 en los editoriales (4%). El siguiente gráfico exhibe los últimos resultados referidos.

Gráfico 7: Resumen categoría tipo de evaluador



8. CONCLUSIONES

8.1 En la presente investigación, se ha dado cuenta, primero, de las diferentes construcciones sintácticas, tanto oracionales como clausulares, empleadas en la expresión de significados evaluativos en distintos géneros de textos expositivos científicos escritos en inglés. Además, se ha determinado, cualitativa y cuantitativamente, qué categorías de significados evaluativos son predominantes en las clases textuales referidas.

Desde los inicios de este trabajo, la identificación de las construcciones sintácticas en los textos examinados puso de manifiesto la insuficiencia descriptiva del modelo formal propuesto por Hunston y Sinclair (2001). Por tanto, fue necesario proceder a una ampliación y reformulación parcial de la taxonomía de los patrones léxico-gramaticales elaborados por los autores referidos. Consecuentemente, con el fin de dar una cuenta

acabada de los diversos patrones sintácticos clausulares empleados en los textos analizados, el modelo descriptivo elaborado ha incluido, finalmente, 53 construcciones sintácticas. Este hallazgo establece un claro contraste con los 6 patrones de construcciones sintácticas evaluativas originalmente establecidos por Hunston y Sinclair (2001).

8.2. Por otra parte, para determinar qué categorías de significados evaluativos son predominantemente empleadas en textos científicos, fue también necesario aplicar, al menos en parte, algunos criterios descriptivos que implican modificaciones parciales de la teoría de la valoración de Martin y White (2005). Primero, dado los objetivos del estudio, se excluyeron los otros dos componentes centrales del modelo general de valoración, a saber, gradación y compromiso. Además, se realizó una modificación del componente actitudinal. Se consideró necesario introducir la categoría 'valoración técnica', propuesta por Soto y Zenteno (2010). Esto pese a que Martin y White (2005) incluyen la subcategoría 'capacidad' como parte de la categoría mayor 'juicio' (definida la primera, en términos amplios, como "…cuán capaces son las personas" (Martin 2001:156), mientras que la segunda es descrita como "…la institucionalización del sentimiento, en el contexto de propuestas (normas sobre como las personas deberían comportarse)" (Martin 2001:155).

La decisión de emplear la categoría 'valoración técnica' en lugar de 'capacidad' (subcategoría no suficientemente explicitada como parte del componente 'juicio') se explica por el hecho de que ella permite una cuenta amplia de no solamente la capacidad de las personas sino también de la eficiencia de su gestión o de la calidad de los procesos y productos derivados de la acción humana. Lo anterior es consecuente con el hecho de que un propósito central del discurso expositivo es comunicar los procesos, productos y resultados de la investigación científica. Por otra parte, si se origina la comunicación de significados evaluativos en géneros discursivos como los aquí estudiados, éstos tienden a

centrarse en la relativa calidad de los hallazgos y productos, no en los ejecutores científicos ni en principios estéticos (criterio propuesto por Martin y White (2005) para establecer la categoría ‘apreciación’). Así lo demuestran, en el primer caso, las 119 instancias de la subcategoría valoración técnica identificadas en el corpus y, en el segundo, la total ausencia de ocurrencias del componente ‘apreciación’ en los textos bajo estudio.

8.3 A pesar de que el volumen de los datos examinados no permite establecer conclusiones definitivas, los resultados cuantitativos obtenidos del análisis permiten responder las 2 preguntas de investigación planteadas en el inicio de nuestro trabajo. En ellas se establecen como interrogantes, primero, qué patrones sintácticos evaluativos presentes en textos expositivos científicos escritos en inglés (según las propuestas teórico-descriptivas de Hunston y Sinclair 2001), demuestran la mayor frecuencia de ocurrencia y, segundo, qué significados valorativos presentes en textos expositivos científicos escritos en inglés (según las propuestas teórico-descriptivas de Martin y White 2005), demuestran la mayor frecuencia de ocurrencia. Asimismo, se ha validado la hipótesis central de nuestro trabajo. En ella, se argumenta que los significados evaluativos (mayoritariamente, de carácter subjetivo) también se encuentran presentes en géneros textuales representativos del discurso expositivo científico escrito en inglés, característica que también comparten los géneros textuales propios del discurso argumentativo y narrativo. Prueba de ello constituyen los 53 patrones sintácticos evaluativos que suman un total de 139 instancias (subdivididas a su vez en 127 propias de la taxonomía elaborada por Prieto y Zenteno para el presente estudio y 12 del modelo general de Hunston y Sinclair 2001). De igual forma, corroboran esta afirmación las 143 instancias de expresiones valorativas identificadas (que se desglosan en los artículos de investigación científica, los artículos periodísticos y los editoriales en 55, 49 y 39 ocurrencias, respectivamente).

8.4 La teoría de la valoración y sus subcomponentes actitudinales (Martin y White 2005) constituye un amplio conjunto de recursos léxico-gramaticales empleados en la construcción de emociones, juicios y valores. La aplicación de este modelo descriptivo en este trabajo permitió dar cuenta, especialmente, de los significados de afecto identificados en el corpus. No obstante, como ya se ha explicado, no facilitó un análisis preciso de, principalmente, los significados expresados acerca de los referentes relacionados con 'los ejecutores de las acciones' o 'los procesos implicados'. Como complemento a esta teoría, el análisis semántico-pragmático demostró la capacidad descriptiva de la reformulación propuesta por Soto y Zenteno (2010).

8.5 En los géneros textuales estudiados (i.e., textos científicos, textos periodísticos y editoriales escritos acerca de un tópico central común, cual es la elaboración del mapa del genoma humano), se ha observado la expresión de evaluaciones acerca del conjunto de los principales referentes del discurso. A la luz de los resultados, parece válido afirmar la existencia de un componente evaluativo en textos científicos, este último frecuentemente caracterizado como "…objetivo e impersonal" (Hunston 1994:192), que pone de manifiesto la subjetividad del enunciador en la comunicación de los resultados a la comunidad científica.

Por cuanto los significados evaluativos, con excepción de aquéllos correspondientes a la categoría valoración técnica, implican la expresión de actitudes y puntos de vista subjetivos es posible plantear, entonces, que la comunicación acerca de referentes propios del dominio científico también implica, por parte del enunciador, la expresión de significados subjetivos como aquellos propios de las categorías 'afecto' y 'juicio' propuestos por Martin y White (2005). Los enunciados en (17) y (18) ilustran este punto.

(17) [AFE (+)] *Interestingly, the chromosomes that have greater than expected number of TAG forests tend to have less than expected number of TAGs deserts*[42].

(18) [JUI (-)] […] *it is unethical for a country to push families towards screening* […][43]

8.6 Los resultados obtenidos en este estudio son coincidentes, en cierta medida, con los hallazgos de otros investigadores del discurso expositivo científico. Al respecto, puede mencionarse a Harvey (2005), quien afirma que tanto en los artículos de investigación como en los artículos de divulgación se evalúa, de preferencia, "…la identidad del objeto de estudio (o de la actividad realizada) y la pertinencia de la investigación, el grado de precisión de los datos y de suficiencia de la información recogida, el beneficio de los resultados alcanzados y/o la probabilidad de que las proyecciones que de los datos se desprenden se cumplan" (2005:99). El siguiente ejemplo caracteriza algunas de las afirmaciones referidas (las expresiones evaluativas en subrayado).

(19) [a] *The availability of SNP data from many different taxa now makes it* [b] *feasible* [c] *to develop a more detailed knowledge of factors* […][44].

En el enunciado (19), extraído de la sección Introducción de un artículo de investigación científica, se observa que, en [a], el grupo nominal expresa, implícitamente, el valor positivo del objeto de estudio y la disponibilidad en el acceso a los recursos. En [b], el grupo adjetivo presenta una evaluación positiva de las acciones y procedimientos necesarios que permitan la solución del problema identificado en la cláusula infinitiva [c].

8.7 En una aproximación comparativa, es posible observar un mayor uso de instancias evaluativas en los artículos de investigación científica (53), número que decrece,

[42] T.10.1.2, lín. 84.

[43] T.10.3.3, lín.54.

[44] T.10.1.5, lín.10.

notoriamente, en los restantes componentes del corpus, i.e., artículos periodísticos y editoriales, con 46 y 40 ocurrencias, respectivamente. Esta tendencia, claramente opuesta a la expresión primaria de la comunicación científica, i.e., la expresión objetiva del conocimiento científico (Bunge 1963), también se manifiesta en los artículos periodísticos, donde es posible observar la expresión de significados de valoración técnica vía metáfora:

(20) *Proteins are the beams and rafters of the cell and the glue that binds the body together; They're the hormones that course through our veins and the guided missiles that target infections; they're the enzymes that build up and break down our energy reserves and the circuits that power movement and thought*[45].

(21) *The flurry of private activity raises the specter of intellectual-property disputes like those that plagued the Human Genome Project*[46].

En los textos editoriales, la situación parece no ser tan distinta a lo expresado por Bolivar (1997), quién señala que éste "…por convención, tiene la función social de evaluar los acontecimientos y los estados de cosas en el mundo" (1997:9)[47]. Asi parece demostrarlo los 12 patrones sintácticos que totalizan un número de 40 expresiones evaluativas identificadas en este tipo de corpus.

8.8 En el desarrollo de esta investigación, realizada en la perspectiva de la lingüística sistémico-funcional, se ha intentado llevar a cabo un estudio comparativo de, primero, los patrones sintácticos empleados en la expresión de significados evaluativos y, segundo, las categorías de significados evaluativos, todos ellos en textos expositivos científicos escritos en inglés. A partir de los resultados y hallazgos, es posible señalar que una aproximación

[45] T.10.2.5, lín. 8.

[46] T.10.2.5, lín. 46.

[47] Bolivar (1997) continúa esta descripción señalando que "…al mismo tiempo, el editorial se caracteriza porque mantiene mejor que ningún otro tipo de texto escrito el carácter dialogal de la interacción natural (Fowler 1991; Bolivar 1994a), lo que obliga al lector a recurrir a esquemas de interpretación ya fijados en la conversación cotidiana".

teórica-analítica a la expresión del componente evaluativo en textos científicos, basada en el modelo descriptivo preliminar aquí planteado, podría ofrecer ciertas ventajas sobre los enfoques originalmente propuestos en el inicio de este estudio. Tanto el uso de los modelos referidos como la posterior modificación de algunas categorías descriptivas permitirían establecer, en primer término, ciertas bases descriptivas validas y aplicables en futuros análisis comparativos de distintos modos y géneros discursivos y, en segundo lugar, otros nichos de investigación del objeto de estudio aquí examinado en el dominio del análisis del discurso, la lingüística textual o la pragmática.

9. REFERENCIAS BIBLIOGRÁFICAS

Barrenechea, Ana M. (1979). Operadores pragmáticos de actitud oracional: los adverbios en –mente y otros signos". En Ana M Barrenechea et al., *Estudios lingüísticos y dialectológicos*. Buenos Aires: Hachette.

Biber, D., S. Johansson, G. Leech, S. Conrad y E. Finegan, (1999). *The Longman grammar of spoken and written English*. Londres: Longman.

Bolivar, A. (1997). La negociación de la evaluación en editoriales de periódicos. *Boletín de Filología* XXXVI:7-24.

Bolívar, A. (2005). Dialogue and confrontation in Venezuelan political interaction. *AILA Review* 18:3-7.

Calsamiglia, H. y A. Tusón. (1999). *Las cosas del decir*. Barcelona: Ariel.

Cassany, D., Luna, M.y G. Sanz (1994). *Enseñar lengua*. Barcelona: Graó.

Channell, J. (2001). Corpus-based analysis of evaluative lexis. En Hunston S. y G. Thompson. (Eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Ciapuscio, G. (1994). *Tipologías textuales*. Buenos Aires: Universidad de Buenos Aires.

Eggins, S. (2004). *An introduction to systemic functional linguistics*. 2a edición. Londres: Continuum.

Eggins, S. y D. Slade. (1997). *Analyzing casual conversation*. Londres: Cassell.

Ferrari, L. (1999). Modalidad y evaluación en artículos de investigación. En Harvey, A. (Ed.), *Actas del III Coloquio Internacional de analistas del discurso (ALED)*. Santiago de Chile: Universidad de Chile y Pontificia Universidad Católica de Chile [En línea]. Disponible en: http://www.fl.unc.edu.ar [Consulta 28/09/2009]

García Negroni, M. (2008). Subjetividad y discurso científico-académico. Acerca de algunas manifestaciones de la subjetividad en el artículo de investigación en español. *Revista Signos* 41(66):5-31.

Halliday, M.A.K. (1985). *An introduction to functional grammar*. 1a edición. Londres: Edward Arnold.

Halliday, M.A.K. (1994). *An introduction to functional grammar*. 2a edición. Londres: Edward Arnold.

Halliday, M.A.K. y Martin, J.R. (1993). *Writing science. Literacy and discursive power*. Londres: The Falmer Press.

Halliday, M.A.K. y Ch. Mathiessen. (2004). *An introduction to functional grammar.* 3a edición. Londres: Edward Arnold.

Harvey, A. (2005). Manifestaciones evaluativas en la ciencia como discurso. En A. Harvey (Comp.), *En torno al discurso. Contribuciones de America Latina*. Santiago de Chile: Ediciones Universidad Católica de Chile.

Hood, S. y J.R. Martin (2005). Invocación de actitudes: el juego de la gradación de la valoración en el discurso. *Revista Signos* 38(58):195-220.

Hunston, S. (1994). Evaluation and organization in a sample of written academic discourse. En M. Coulthard (Ed.), *Advances in written text analysis.* Londres: Routledge.

Hunston, S. y G. Thompson (Eds.). (2001). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Hunston, S. y J. Sinclair. (2001). A local grammar of evaluation. En Hunston S. y G. Thompson. (Eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Hyland, K. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction* 1(1):5-22.

Jaffe, A. (Ed.) (2009). *Sociolinguistic perspectives*. Nueva York: OUP.

Johansson, M. (2001). Discurso e ideología en la problemática de las violaciones a los derechos humanos: un análisis crítico de editoriales sobre el informe Rettig. *Tesis para optar al grado de Magíster en Lingüística con mención en Lengua Hispánica*. Santiago de Chile: Universidad de Chile.

Kerbrat-Orecchioni, C. (1993). *La enunciación de la subjetividad en el lenguaje*. Buenos Aires: Edicial.

Labov, W. (1972). *Language in the inner city: studies in the black English vernacular*. Filadelfia: University of Pennsylvania Press.

Labov, W. y J. Waletzky (1967). Narrative analysis: Oral versions of personal experience. En J. Helm (Ed.), *Essays on the verbal and visual arts*. Seattle, Washington: University of Washington Press.

Leech, G., Deuchar, M. y R. Hoogenraad (2006). *English grammar for today: a new introduction*. 2a edición. Basingstoke, Hampshire: Palgrave MacMillan.

Martin, J.R. (2000a). Beyond exchange. Appraisal systems in English. En Hunston S. y G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Martin, J.R. y D. Rose. (2003). *Working with discourse: meaning beyond the clause.* Londres: Continuum.

Martin, J.R. y P.R.R. White. (2005). *The language of evaluation. Appraisal in English*. Basingstoke, Hampshire: Palgrave Mcmillan.

Macken-Horarik, M. y J.R. Martin (Eds.) (2003). Appraisal. Volumen especial. *Text* 23(2).

Myers, G. (1994). Narratives of science and nature in popularizing molecular genetics. En M. Coulthard (Ed.), *Advances in written text analysis*.Londres: Routledge.

Ochs, E. y B. Schieffelin. (1989). Language has a heart. *Text* 9(1):7-25.

Quirk, R., S. Greenbaum, G. Leech y J. Svartvik. (1985). *A comprehensive grammar of the English language*. Londres: Longman.

Shiro, M. (1999). Representación del yo y del otro en la narrativa de adultos caraqueños. En Harvey, A. (Ed.), *Actas del III Coloquio Internacional de analistas del discurso (ALED)*. Santiago de Chile: Universidad de Chile y Pontificia Universidad Católica de Chile [En línea]. Disponible en: http://www. filosofia.uchile.cl [Consulta 24/07/2007].

Shiro, M. (2004). Expressions of epistemic modality and the construction of Narrative stance in Venezuelan children's stories. *Psychology of language and communication*. 8(2):35-56.

Smith, C. (1997). *The parameter of aspect*. Dordrecht: Kluwer.

Soto, G. y C. Zenteno (2001-2003). La subtopicalización en el discurso científico escrito. *Lenguas Modernas* 28-29:29-52.

Soto, G. y C. Zenteno. (2010). La relación en el discurso entre el pretérito perfecto compuesto y las expresiones evaluativas. Análisis de una cuenta pública presidencial. *Revista Signos* 43(73):333-354.

Swales, J. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research Genres. Explorations and applications*. Cambridge: Cambridge University Press.

Tadros, A. (1994). Predictive categories in expository text. En M. Coulthard (Ed.), *Advances in written text analysis*. Londres: Routledge.

White, P. (2001). An introductory tour through appraisal theory. [En línea]. Disponible en: http://www.grammatics.com [Consulta 20/05/2007]

Yancovic, B. (2007). *El genoma humano al alcance de todos.* Santiago de Chile: RIL editores.

Zenteno, C. (1996). Algunas propuestas para una gramática textual en la enseñanza del inglés como lengua extranjera. *Lenguas Modernas* 23:167-180.

Zenteno, C. (1997). Hacia una gramática textual en la enseñanza del inglés como lengua extranjera. *Lenguas Modernas* 24:169-185.

Zenteno, C. (2002-2003). La opinión y la evaluación en un editorial de prensa. *Boletín de Filología* XXXIX:305-324.

9.1 Artículos de investigación cientìfica

Constantini, M., Clay, O., Auletta y G. Bernardi (2006). An isochore map o human chromosomes. *Genome Research* 16:536-541. [En línea]. Disponible en http:// www.genome.org [Consulta 03/02/2009].

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G. Pačes, J., Burt A. y M. Tristem (2004). Long-term reinfection of the human genome by endogenous retroviruses. *Proceedings of the National Academy of Sciences of the United States of America*. [En línea]. Disponible en http://www.pnas.org [Consulta 04/02/2009].

D'Onofrio, E. (2002). Expression patterns and gene distribution in the human genome. *Gene* 300:155-160. [En línea]. Disponible en http://nslij-genetics.org [Consulta 03/02/2009].

Morton, B., Bi, I., McMullen, M. y B. Gaut. (2006). Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. *Genetics* 172:569-577. [En línea]. Disponible en http://www.genetics.org [Consulta 04/02/2009].

Shoja, V. (2006). A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat. *Molecular Biology and Evolution* 23(11): 2134–2141.[En línea]. Disponible en http://mbe.oxfordjournals.org [Consulta 05/02/2009].

9.2 Artículos periodísticos

Duncan, D. (2005). Researchers focus on differences between groups to find bad DNA. *Discover Magazine*. [En línea]. Disponible en http://discovermagazine.com [Consulta 05/02/2009].

Kher, U. (2000). Beyond genomics: the next frontier: proteomics. *Time Magazine* [En línea]. Disponible en http://www.time.com [Consulta 05/02/2009].

Lovgren, S. (2007). Macaque genome deciphered, may herald medical breakthroughs. *National Geographic News*. [En línea]. Disponible en http:// news.nationalgeographic.com [Consulta 07/02/2009].

Minkel, J.R. (2007). The 1 percent genome solution. *Scientific American*. [En línea]. Disponible en http://www.scientificamerican.com [Consulta 07/02/2009].

Sanders, L. (2008). First complete cancer genome sequenced. *Science News*. [En línea]. Disponible en http://www.sciencenews.org [Consulta 06/02/2009].

9.3 Editoriales

Balaram, P. (2000). Post-genome blues. *Current Science* 79(1):2-3. [En línea]. Disponible en http://www.ias.ac.in [Consulta 07/02/2009].

Dhand, R. (2006). The 'finished' landscape. *Human Genome Collection* 1:7. [En línea]. Disponible en http://www.nature.com [Consulta 07/02/2009]

Kennedy, D. (2003). DNA: one teacher's reflection. *Science* 300(5617):213. [En línea]. Disponible en http://sciencemag.org [Consulta 06/02/2009].

Nature. (2004). Human genomes, public and private. *Nature* 409: 745. [En línea]. Disponible http://www.nature.com [Consulta 07/02/2009].

Williamson, B. (2001). Our human genome –how can it serve us well? *Bulletin of the World Health Organization* 79(11):1005. [En línea]. Disponible en http://www.scielosp.org [Consulta 07/02/2009].

10. APÉNDICE

10.1 Artículos de investigación científica

10.1.1 An isochore map of human chromosomes

Introduction

Well before genome sequencing, ultracentrifugation in Cs_2SO_4 density gradients in the presence of sequence-specific ligands (e.g., Ag^+) was shown to lead to a high resolution of mammalian DNAs according to base composition (Corneo *et al.* 1968). These findings opened a new inroad in the study of the organization of eukaryotic genomes, superseding DNA reassociation kinetics (Britten and Kohne 1968), which was based on the separation of single-and double-stranded DNA on hydroxyapatite (Bernardi 1965). The new density gradient approach showed that, neglecting satellite DNAs, the genomes of warm-blooded vertebrates were characterized by a striking long-range compositional heterogeneity (Filipski *et al.* 1973; Macaya *et al.* 1976; Thiery *et al.* 1976). Indeed, these genomes are mosaics of isochores, long (>300 kb), compositionally fairly homogeneous regions that belong to a small number of families characterized by different average GC levels (Macaya *et al*. 1976), and are associated with basic biological properties (for reviews, see Bernardi *et al.* 1985; Bernardi 1995, 2004).

A quarter of a century after the original studies that had defined the approximate sizes and compositions of isochores as well as the compositions and relative amounts of isochore families, it was reported that isochores could not be identified in the draft sequence of the human genome (Lander *et al.* 2001), starting a debate that is still ongoing. The different computational approaches used to disprove or redefine isochores (Eyre-Walker and Hurst 2001; Häring and Kypr 2001; Lander *et al.* 2001; Nekrutenko and Li 2001; Cohen *et al.* 2005) were, however, shown to be inadequate (Bernardi 2001; Clay and Bernardi 2001a, b, 2005; Li 2002; Oliver *et al.* 2002, 2004; Li *et al.* 2003; Melodelima *et al.* 2005), even if some of them led to a partial identification of isochores. This debate prompted us to map the isochores, as originally defined (Macaya *et al.* 1976), in the finished sequence of the human genome (International Human Genome Sequencing Consortium 2004).

Average GC levels were, therefore, assessed over long DNA stretches (>200 kb), while GC variation was estimated by measuring standard deviations of GC over such stretches using a 100-kb moving window. The findings reported here are in agreement with previous results obtained by equilibrium sedimentation and confirm the existence of five isochore families (see below), but they go much farther in that they directly identify and map isochores on chromosomes, thus leading to a resolution of >3000 chromosomal bands.

Almost 50 years ago, calf thymus DNA, the standard eukaryotic DNA, was shown to be remarkably more heterogeneous in base composition than bacterial DNAs (Meselson *et al.* 1957). In fact, the very strong heterogeneity was largely, but not entirely, due to the presence of GC-rich satellites that represent 23% of the bovine genome (see Bernardi 2004). Interestingly, high-resolution ultracentrifugation of bovine DNA was not only able to separate the GC-rich satellites, but also showed a discontinuous compositional heterogeneity of the main band (Filipski *et al.* 1973), consisting of three families of DNA molecules. This was in contrast with the then predominant view (still defended by some authors) (Galtier *et al.* 2002) of a continuous heterogeneity of main-band DNA. The families of DNA molecules were then shown to be present in the other mammalian genomes explored and to derive from longer, fairly homogeneous DNA stretches (Macaya *et al.* 1976; Thiery *et al.* 1976) that were called isochores (Cuny *et al.* 1981) for (compositionally) equal landscapes. Later work resolved the first family into two families, L1 and L2, named the second and the third families H1 and H2, respectively, and identified another quantitatively small family, H3 (Zerial *et al.* 1986). Most mammalian genomes, including the human genome, are made up of five families of isochores -L1, L2, H1, H2, and H3- in order of increasing GC levels, if satellite DNAs (2% of the genome) and ribosomal DNAs (0.5% of the genome) are neglected. Incidentally, satellite and ribosomal DNAs may also be considered isochores, because of their compositional homogeneity (Bernardi 1995).

Isochores were shown to be tightly linked to basic biological properties, such as gene density, replication timing, and recombination (see Bernardi 2004), and have, therefore, been considered "a fundamental level of genome organization" (Eyre-Walker and Hurst 2001). In order to characterize isochores, the present work did not take into account their biological properties (as done by Melodelima *et al.* 2005) but only relied on two parameters, GC levels and their standard deviations.

Discussion

The present findings, while confirming the isochore features previously established, push our knowledge farther, by quantifying the size, GC levels, standard deviations, and coordinates of isochores on the human genome map. Moreover, these findings also indicate that isochores may be visualized as the ultimate banding patterns of the chromosomes in warm-blooded vertebrates, and that they are arranged in blocks, corresponding to chromosomal bands at the standard 850 band resolution.

It seems appropriate here to briefly summarize two major points of interest concerning isochores. From a practical viewpoint, isochores allowed us to gain an insight into the genome organization of warm-blooded vertebrates (and of other organisms) (Bernardi 2004). Localizing genes in separate isochores led to the discovery of an unexpected and strikingly nonrandom distribution of genes (Mouchiroud *et al.*1991; Zoubak *et al.*1996), which were found in two "gene spaces." The "genome core," composed of the isochore families H2 and H3, comprises more than half of the genes even though they represent only 15% of the genome, whereas the "genome desert" (the isochore families L1, L2, and H1) is made up of large expanses with low and often extremely low gene densities. These two gene spaces are characterized by several different properties (for review, see Bernardi 2004), the most remarkable ones being the correlations of isochore families not only with gene density but also with replication timing, recombination, and location and chromatin structure in interphase nuclei, chromatin being "open" in the genome core and "closed" in the genome desert (Saccone *et al.* 2002).

From a more general point of view, the present results raise one major question concerning the origin and maintenance of GC-rich isochores, which are a common, characteristic property of the genomes of warm-blooded vertebrates. We now know that the GC-rich (and gene-rich) isochores are the result of GC increases in the corresponding gene-rich regions of cold-blooded vertebrates, which are much less GC rich (see Bernardi 2004). We proposed that the increasing body temperature accompanying the emergence of homeothermy led to a need for a thermodynamic stabilization of DNA (Bernardi and Bernardi 1986). At the transition between cold and warm-blooded vertebrates, GC-poor isochores did not undergo any significant compositional change because they were stabilized by their closed chromatin structures, whereas GC increases took place in the gene-rich genome regions that were characterized by an open chromatin structure (Federico *et al.* 2006). In fact, the stabilization also concerned RNA and proteins (GC-rich codons favoring amino acids that lead to a higher thermal stability) (Bernardi and Bernardi 1986; Nishio *et al.* 2003). Although several explanations have been proposed for the GC increases, which took place in spite of a strong AT bias of the mutation process (Gojobori *et al.* 1982; Smith and Eyre-Walker 2001; Alvarez-Valin *et al.* 2002; Santini and Bernardi 2005), the only one that was compatible with all known facts was natural selection, mainly negative selection (for review, see Bernardi 2004). How this can be reconciled with the fact that the vast majority of mutations are neutral was explained by the neo-selectionist theory, which posits that neutral GC→AT changes (the AT bias) are tolerated until they cause a regional compositional change in DNA. In turn, this can change chromatin structure and the interaction with regulatory proteins, thus impairing the correct gene expression and leading to negative selection (Bernardi 2004).

10.1.2 A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat

Introduction

DNA duplication is the principle process by which the genetic raw material is provided for the origin of evolutionary novelties such as new gene function and expression patterns and is important in adaptive evolution (Wolfe and Li 2003). Possible duplication mechanisms include unequal crossover (i.e., tandem duplication), retrotransposition, and replicative translocation (Ohno 1970; Nei 1987). Duplication events may vary in content and frequency: whereas whole-genome duplications appear to be limited, small tandem duplications appear to be quite common. In fact, localized duplication of genomic segments and rearrangement of chromosomal segments have been proposed to be 2 major factors in eukaryotic genome evolution (Eichler and Sankoff 2003).

The availability of complete genomic sequences makes it possible to investigate how genomes are structured by different mechanisms of gene duplication. In this paper, we focused on tandem duplication. Tandem duplication of related genes has been shown to act as the driving evolutionary force in the origin and maintenance of gene families (Reams and Neidle 2004) and has been a common mechanism of genetic adaptation to environmental challenges in organisms such as bacteria (Anderson and Roth 1977; Roth *et al.* 1996; Hastings *et al.* 2000), yeast (Brown *et al.* 1998), mosquitoes (Lenormand *et al.* 1998), plants (Harms *et al.* 1992; Shyr *et al.* 1992; Leister 2004), and humans and other mammals (Stark 1993).

Specifically, we identified all tandemly arrayed genes (TAGs) in the genomes of humans, mouse, and rat and addressed the following issues: First, because duplicated genes can be arranged in tandem or dispersed on different chromosomes, we want to determine how many duplicated genes are in tandem arrays. This will shed light on the contribution of tandem duplication to gene duplication in the 3 mammalian genomes. Second, we are interested in examining the chromosomal distribution of TAGs to see whether there is significant clustering of TAGs on some chromosomal regions.

Third, about 70% of the TAGs in the Arabidopsis thaliana genome have only 2 members in the array (Zhang and Gaut 2003), so do the 3 mammalian genomes show a similar pattern? Fourth, is there any nonrandom association between gene function defined by Gene Ontology (GO) categories and TAGs? We expect that genes with certain functions may prefer tandem arrangement over other types of more dispersed spatial arrangements as tandem arrangement can either entail high probability of generating more duplicated copies or promote a desired degree of diversity or homogeneity via concerted evolution (Ohno 1970). Fifth, it has been hypothesized that the preferred orientation of TAGs is parallel because locating on different strands is detrimental to the stability of the array (Graham 1995). We thus examined the orientations of array members and compared them with the genome pattern. Finally, is tandem duplication a preferred mechanism of duplication for large gene families? In other words, do we observe more TAGs in larger families than smaller ones?

Discussion

Significance of tandem duplication

Previous and current studies all suggest that TAGs are a major component of the genome. The percentages of TAGs in different genomes of plants and animals span a narrow range of 10–17% (10% for Caenorhabditis elegans [Semple and Wolfe 1999]; 17% for A. thaliana [Zhang and Gaut 2003]; 14% for Oryza sativa [Yu *et al.* 2005]; 14–17% for humans, mouse, and rat [table 2]). Moreover, about 21–25% of the duplicated genes in the 3 mammalian genomes are TAGs (table 2), suggesting that tandem duplication is a major method of gene duplication in many genomes.

Studies on recent duplication in several mammalian genomes show that intrachromosomal duplications are more common than interchromosomal duplications (Cheung *et al.* 2003; Eichler and Sankoff 2003; Friedman and Hughes 2004; Zhang *et al.* 2005). Intrachromosomal duplication may include one or more genes and depending on the locations and the mechanisms of duplication, it can be tandem duplication. In fact, the number of intrachromosomal duplicated genes is significantly correlated with the number of TAGs for the 3 species (P value 5 13 10-5, see supplementary material online).

Contribution of tandem duplication to different sized gene families

The research on TAGs has been largely confined to individual families of TAGs that serve important physiological functions, such as ribosomal RNA genes, histone genes, immunoglobulin genes, and MHC genes. In these large gene families, most of the members arose through tandem duplication. The unanswered question remains as to whether tandem duplication is a more favored duplication mechanism in large gene families than small ones. It is expected that the larger the family is, the more likely the family resorts to tandem duplication as an efficient way of creating more duplicated genes (Ohno 1970).

To examine this issue, we calculated the average proportion of TAGs in gene families of different sizes for all 3 genomes. The average proportion of TAGs in gene families of different sizes ranges from 15% to 33% in 3 genomes and appears to be higher in large families than small families (table 5). However, although the average proportions of TAGs and family sizes are positively correlated (i.e., large families tend to have on average more TAGs than smaller ones) in humans (Spearman's rank correlation coefficient q 5 0.91, P value 5 0.00047) and mouse (q 5 0.71, P value 5 0.0275), but not in rat (q 5 0.53, P value 5 0.1133), the proportion of TAGs in all individual families and family sizes do not show significant correlation (P value .0.05). The observation that large families tend to have on average higher percentages of TAGs than small ones could be due to the possibility that large families have a high likelihood of being tandem through random arrangement. However, our simulation (figure 2 and supplementary material online) shows that this is unlikely because random permutations of all the duplicated genes yield a tiny amount of nonreal TAGs, and for large gene families, which is a much smaller subset of all duplicated genes, the proportions of nonreal TAGs should be even smaller.

Distribution of TAGs on the chromosomes

In all species, TAG distribution shows great heterogeneity along the chromosomes with some chromosomes enriched with TAGs and some depleted of TAGs (fig. 4). Using the definitions of TAG deserts and forests, we studied TAG enrichment and depletion with respect to individual chromosomes. Interestingly, the chromosomes that have greater than expected numbers of TAG forests tend to have less than expected numbers of TAG deserts (fig. 5 and Supplementary Material online), suggesting that TAGs have preferences for chromosomes. Furthermore, using the information on subtelomeric and pericentromeric regions in humans, we found that TAGs tend to be enriched in pericentromeric regions and thus have preference for specific locations as well. In this regard, it is worth noting that it has been shown that pericentromeric regions are enriched with recent segmental duplications in humans (Bailey *et al.* 2001; Zhang *et al.* 2005). As mentioned before, some of the recent segmental duplications in humans might be in fact tandem duplication. Therefore, the common preference for pericentromeric regions by tandem duplication and recent segmental duplication might not be a coincidence.

An interesting question relevant to the TAG distribution is whether the regions that are enriched with TAGs are also rich in other non-TAG duplicates. Our analysis shows that in humans, the 2 regions that are statistically enriched in TAGs also show enrichment of other non-TAG duplicates, whereas in rat, the regions that are rich in TAGs show depletion of non-TAG duplicates (results not shown).

Distribution of TAG sizes

It has been observed that the majority of TAGs have only 2 members in the array in many genomes such as A. thaliana (Zhang and Gaut 2003), C. elegans (Semple and Wolfe 1999), and rice (Yu *et al.* 2005), suggesting that it might be a rather general phenomenon in eukaryotes (fig. 3). The distribution of TAG sizes can be described by a power law distribution, a common type of distribution that appears in various biological quantities such as the distribution of gene family sizes in different eukaryotes (Enright *et al.* 2002) and prokaryotes (Huynen and van Nimwegen 1998).

However, for the majority of genes, diversity in function might be more preferred than quantity, as clearly demonstrated by genes with binding and receptor activities and disease resistance functions (table 4).

TAG orientations and intergenic distances

Graham defined ‘‘tandem arrays’’ as arrays in which a DNA segment is repeated head to tail, with all copies in the same orientation, and suggested that unequal recombination homogenizes head-to-tail tandem arrays but would cause arrays with oppositely oriented repeats to undergo disastrous duplication–deletion events, which results in these arrays being rare (Graham 1995). The definition of TAGs in the current study is somewhat different from that of Graham’s. Nevertheless, the results show that compared with the genome, parallel orientation in TAGs appears to be more favored than divergent or convergent orientations (table 3), corroborating Graham’s conjecture, at least in the 3 genomes. Furthermore, consistent with the great disparity between the proportion of parallel orientations in TAGs and that in the genomes, the intergenic distances of genes in parallel orientations also show the greatest disparity between TAGs and the genomes compared with the distances among genes in convergent and divergent orientations (see supplementary material online). This raises the question of the evolutionary significance of parallel orientations in TAGs: why TAGs with parallel orientation show distinct patterns from that in the genome. Is there any adaptive significance with parallel orientation or does the observed pattern simply reflect the pattern of unequal crossover? To our knowledge, there have been no previous studies examining the effect of parallel versus other types of orientations in TAGs other than Graham’s hypothesis. More studies are needed to investigate the underlying mechanism and the nature of this phenomenon.

10.1.3 Expression patterns and gene distribution in the human genome

Introduction

The localization of 40 human genes in isochore families first showed that genes were not uniformly distributed in the human genome, being more concentrated in GC-rich isochores (Bernardi *et al.,* 1985). Thereafter, *in silico* localization of ~1400 human genes (D' Onofrio *et al.*, 1991) led to the same conclusions (Mouchiroud *et al.*, 1991), further confirmed by larger sets of genes (Zoubak *et al.*, 1996; Saccone *et al.*, 2001).

The biased gene distribution in the human genome raised a question about the correlation between gene distribution and gene expression patterns, or in other words, about the distribution of tissue-specific and widely expressed genes according to GC levels of the isochores.

The first attempt to answer the above question, summarizing independent experimental results on chromatin structure and gene composition, as well as gene distribution, led up to the hypothesis that: 'the H3 isochore family presumably has the highest level of transcription because of its very high concentration of genes - especially housekeeping genes' (Bernardi, 1993; and references therein). The high expression levels of the genes localized in H3 were further supported by *in silico* investigation on the sequence context of the AUG start codon. The results showed that genes located in the GC-rich isochores required highly efficient translation (Pesole *et al.*, 1999).

Subsequent analyses on the correlation between gene distribution and gene expression levels showed that the majority of the widely expressed genes were localized mainly in GC-poor isochores, whereas tissue-specific genes were localized in the GC-rich ones (Gonçalves *et al.*, 2000). The authors drew these conclusions by analyzing the base composition and the distribution of genes with or without retropseudogenes, the former being more widely expressed than the latter. However, in the human genome, using a different algorithm, the propensity for retro transposition was found to be unaffected by the GC content of the genes (Venter *et al.*, 2000).

Studying the origin of CpG islands, tissue-specific genes were confirmed to be mainly localized in GC-rich isochores (64% in H3), whereas widely expressed genes distribution was independent of the isochore content (Ponger *et al.*, 2001). The finding led Galtier *et al.* (2001) to argue: 'selection, if any, must be unrelated to the gene expression level or pattern'.

However, in both paper dealing with the correlation between gene distribution and expression patterns (Gonçalves *et al.*, 2000; Ponger *et al.*, 2001), the gene partition was performed according to the criteria defined in Mouchiroud *et al.*, (1991). [In the last decade, however, several results based on theoretical and experimental approaches led to an improvement of the gene partition criteria] (Saccone *et al.*, 1993, 1996, 1999; Zoubak *et al.*, 1996; Federico *et al*., 2000).

In the present paper the distribution of widely expressed genes in the human genome was revisited by analysing the dataset of human genes collected in the CpG islands database (Larsen *et al.*, 1992; database 4.0, 1996), as well as a dataset of human genes orthologous to those of Xenopus, calf and murids. The results from the two independent datasets led to the conclusion that widely expressed genes: (i) are mainly localized in GC-rich isochores; (ii) are not the majority of the genes in the GC-rich isochores; and (iii) are not GC3 poorer than tissue-specific genes.

Conclusions

Re-examination of experimental and theoretical data from publications spanning almost 10 years (Saccone *et al.*, 1993, 1996, 1999; Zoubak *et al.*, 1996: Federico *et al.*, 2000) allowed us to refine the gene partition criteria first used by Mouchiroud *et al.*, (1991). Two independent datasets were analyzed: one was the updated Larsen's database (Larsen *et al.*, 1992, database 4.0, 1996). The other a set of available human genes orthologous to genes from Xenopus, calf and murids. The analysis of the gene frequencies in GC-poor and GC-rich regions led us to reach the conclusions that tissue-specific and widely expressed genes:

- show no compositional differences at GC3 level;
- are in a ratio higher than one in GC-rich isochores, that is in these isochores tissue-specific genes are more abundant than widely expressed ones;

- follow the general gene distribution, reaching the highest frequency in the GC-rich isochores.

The last point deserves some comments. The gene density and the GC level of isochores were found to be correlated, and both were hypothesized to be correlated with the gene expression levels (Bernardi, 1993). The hypothesis was one of the key points supporting that the human genome organization and evolution are under selective forces (Bernardi, 2000a, for a review). Gultier *et al.* (2001) rejected the hypothesis arguing "such a selective pressure without apparent correlation with gene expression appeared quite speculative". The statement was based essentially on the observation that "the GC content of the genes does not correlate positively with their expression level or pattern (Gonçalves *et al*., 2000)". In contrast, our present results support the view that the GC level of the genes is correlated with their expression patterns; indeed both tissue-specific and widely expressed genes are concentrated in the GC-rich isochores. Can we answer the second question, whether a correlation between the GC content and the expression levels of the genes also holds?

An analysis of gene transcription levels revealed 'a high order organization of the genome' (Caron *et al.,* 2001). Indeed, regions of increased gene expression (RIDGE) were also characterized by high gene density, a correlation found for 50-60% of the RIDGEs (Caron *et al.,* 2001). Unexpectedly, it was also reported, 'about 40-50% of RIDGEs are not gene dense. These RIDGEs preferentially maps to telomeres' (Caron et al., 2001). Experimental results on human metaphase chromosomes showed the telomeric localization of H3, the isochore family with the highest gene concentration and GC content (Saccone *et al.,* 1999). Unfortunately, the figures representing the RIDGEs distribution (Caron *et al*., 2001) did not allow a precise localization of those regions along the chromosomes. Indeed, chromosomes were represented without correlation with the corresponding banding or length. Therefore, at present, it is difficult to assess the exact correspondence of the telomeric region described in Caron *et al.* (2001), with those described in Saccone *et al.* (1999).

However, it was reported that chromosomes 4, 13, 18 and 21 were completely devoid of RIDGEs showing also “low gene expression and low gene density” (Caron *et al*., 2001). The same chromosomes turned out: (i) to have, on the average, very low GC levels: 38%, 38%, 40% and 41%, respectively (Venter *et al.,* 2001); and (ii) to have low amount of GC-rich isochores, from a compositional profile obtained by a sliding window analysis (100 Mb) (Pavlicek *et al.,* 2002). From there, it could be argued that RIDGEs distribution correlates not only with the gene density but also with the GC content of the isochores.

The original hypothesis of Bernardi (1993) that the GC rich “isochore family presumably has the highest level of transcription because of its high concentration of genes” was fundamentally correct. However, it should be expected that the transcriptional levels affect the gene concentration. Therefore, we can conclude that the force driving the non-uniform gene distribution in the human genome is the expression level of the genes.

10.1.4 Long-term reinfection of the human genome by endogenous retroviruses

Introduction

Endogenous retroviruses (ERVs) represent the proviral phase of exogenous retroviruses that have integrated into the germ line of their host. They typically consist of an internal region with three genes (‘gag’, ‘pol’, and ‘env’) plus two flanking, noncoding LTRs, which are identical at the time of integration. Human ERVs (HERVs) comprise ≈5 -8% of the human genome, with 98,000 elements and fragments, but phylogenetic analysis of conserved regions within their ‘pol’ and ‘env’ genes indicates that they form only a small number of clades among nonhuman exogenous and endogenous retroviruses. Thus, there appears to have been a huge proliferation of elements derived from only a few initial germ-line invasions by exogenous viruses. Over time, replication-competent ERVs accumulate in-frame stop codons and frame-shift mutations as a result of host DNA replication and, within the human genome sequence, these processes have led to the inactivation of almost every element. Another mechanism by which ERVs are inactivated is via recombinational deletion between the two viral LTRs, which removes the internal region leaving a solo LTR structure. Solo LTRs are typically 10-100 times more numerous than their more intact, undeleted counterparts.

Surprisingly, although the processes that inactivate ERVs are known, the mechanism by which HERVs have increased in copy number is only poorly understood, though several candidate mechanisms have been proposed. ERVs could undergo retrotransposition within germ-line cells, and do this via two routes: either in ‘cis’, where the virus uses its own encoded proteins for mobilization, the predominant method in long interspersed nuclear elements (LINEs), or by complementation in ‘trans’, where the proteins are supplied by another endogenous or exogenous virus within the same cell. Retrotransposition in ‘cis’ does not require an intact ‘env’ gene (which is necessary only for movement outside the cell), whereas complementation in ‘trans’ does not require the virus to have any functional genes (merely requiring a promoter and other motifs for expression and packaging of the viral RNA).

ERVs could also increase in copy number by reinfection. Reinfection can occur between germ-line cells or by infection of germ-line cells by viruses originating in somatic cells. Thus, it does not necessarily require the viruses to be passed between different individuals in a population. Studies on murine leukemia virus proviruses in mice and ‘gypsy’ retroelements in ‘Drosophila’ show that new elements can integrate into the germ line by extracellular infection. All these mechanisms can be expected to have left different patterns in the nucleotides of existing ERVs.

Within humans, the most recently active ERVs are members of the HERV-K (HML2) family. This family first integrated into the genome of the common ancestor of humans and Old World monkeys at least 30 million years ago, and it contains >12 elements that have integrated since the divergence of humans and chimpanzees, as well as at least two that are polymorphic among humans. This recent activity makes this family ideal for distinguishing between the alternative mechanisms of proliferation.

Here we show that the HERV-K (HML2) family has increased in copy number predominantly via reinfection, and that the family has probably retained replication-competent and infectious members for >30 million years. We also present evidence for persistent reinfection by other ERV families within the human genome, and suggest that endogenous retrovirus families are often capable of extremely long periods of smoldering infection.

Discussions and conclusions

The paucity of inherited stop codons, and the low d_N/d_S ratios for all genes (including ‘env’) within the internal branches of the HERV-K (HML2) phylogeny, strongly indicates purifying selection, which in turn suggests that this family has increased in copy number predominantly by reinfection rather than by retrotransposition in ‘cis’, or complementation in ‘trans’. Proliferation via the latter two mechanisms would result in both the presence of numerous shared stop codons within the ‘env’ gene and also in the neutral evolution of this gene (with a corresponding d_N/d_S ratio close to one), but this is not the case.

Such reinfection may only involve movement from somatic to germ-line cells within the same individual, and does not necessarily require transfer between different individuals in the host population.

The apparent rarity of complementation in 'trans' may seem surprising given our understanding of the mechanism of retroviral replication. One possible explanation is that often there is only a single element expressed in any particular cell, and so complementation in 'trans' is not usually possible. The apparent rarity of retrotransposition in 'cis' (as opposed to between-cell reinfection) is also surprising, and may suggest that gene expression is more rigidly controlled in germ-line cells than in the surrounding somatic cells.

Previously, Costas demonstrated low d_N/d_S ratios in pairwise comparisons of genes from selected HERV-K (HML2) elements and suggested that there had been some transpositional activity since the human chimpanzee divergence some six million years ago. Here, however, in addition to showing that proliferation is largely caused by reinfection, our analyses allow us to make additional inferences about the evolution of the HERV-K (HML2) family. Continuous selection strongly suggests continuous functionality, and this implies that the HERV-K (HML2) lineage has retained replication-competent members since its origin. Furthermore, because selection is a property of a population that involves replication and loss of unfit elements, a pool of active elements must have been present throughout this period. Thus, the internal branches in our phylogeny do not represent the same element at different times, or single copying events. Rather, they represent samples from a changing pool of endogenous retroviruses (evolving primarily via viral rather than host DNA replication) that have been both replication competent and infectious throughout the evolution of the HERV-K (HML2) family. Most members of this pool would not have been fixed (i.e., they were only present within some individuals of the population) and there would have been a continuous turnover as elements were lost via natural selection and genetic drift, and gained via new germ-line integrations. By contrast, the terminal branches in our phylogeny represent elements that have become both inactive and fixed, presumably by means of genetic drift.

Thus, they account for almost all of the stop codon acquisitions and have a much higher d_N/d_S ratio. Even so, the terminal branches still have a d_N/d_S ratio of <1, possibly because many elements have been lost (via recombinational deletion or genetic drift), which will have merged internal branches onto terminal branches.

Although our phylogeny shows a rapid increase in the number of elements following the origin of humans, and changes shape from wholly pectinate (comb-shaped) to more balanced, this should not be taken as evidence for a human-specific burst of activity. We suggest that the change in tree topology has been caused simply by there having been less time for recent elements to be lost by recombinational deletion (leaving only solo LTRs). In the future, when most of the modern elements have undergone such events, the phylogeny of those remaining is likely to be sparse and pectinate, as the phylogeny of the older elements is now. We found evidence of this process occurring at the present time in the relatively young, human-specific element K103. Although most humans carry the full-length virus, we found that the human genome project sequence and several sub-Saharan African individuals have only a solo LTR at the integration site (unpublished data).

The scenario of continuous purifying selection described above explains why the human genome still contains relatively intact HERV-K (HML2) elements, even after a 30-million-year-long association with this particular viral lineage. Previously, such elements were thought to result from the mutational or recombinational reactivation of elements that were otherwise continually decaying due to errors introduced during host replication. Our examination of 'env' d_N/d_S ratios in other HERV families suggests that the dominant role of reinfection and the persistence of a pool of active elements may be a general feature of HERV evolution. However, we note that one HERV family (HERV-L) does not encode an 'env' gene, and that in another (HERV-H), >90% of the elements share large deletions in 'pol' and 'env'. Thus, in these cases (which are the two largest families in terms of copy number), proliferation has probably occurred via alternative mechanism(s) to reinfection.

Until now, our understanding of the evolution of interspersed repeats such as HERVs has been influenced heavily by phylogenetic tree shape, and the typically unbalanced phylogenies have been thought to reflect one or a few active elements (called masters) giving rise to many other copies that do not copy themselves. In such models, applied also to the HERV-K (HML2) family, the internal nodes represent the same master element (at the same integration site) at different times. In these scenarios, the master elements can remain functional for long periods of time, but they would still be expected to accumulate synonymous and nonsynonymous substitutions at the same rate. Hence, the d_N/d_S ratios on the internal branches of the HERV-K (HML2) phylogeny would not be significantly different from 1. In contrast, our results show that the internal branches are under strong purifying selection, and internal nodes represent different elements that are survivors from a pool of unfixed, active endogenous retroviruses. Thus, although there are no HERV-K (HML2) elements in the completed human genome sequence with a full coding capacity (which has led to the conclusion that this family is unlikely to be capable of causing disease; refs. 7 and 33), our data indicate that, as suggested by previous authors (14, 26), a proportion of the human population may still harbor active and infectious members of the HERV-K (HML2) family.

10.1.5 Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition

Introduction

Evolution is ultimately dependent on mutation and thus characterizing mutation rates and biases, within and among genomes, is a prerequisite for studying genomics and molecular evolution. For example, comparative genomics requires an understanding of mutation dynamics in different lineages (e.g., Dermitzakis *et al.* 2002), and compositional patterns such as the possible isochore structure in vertebrates (Bernardi 2000, but see Cohen *et al.* 2005) cannot be adequately studied without an understanding of how mutation bias varies along chromosomes (*e.g.*, Duret *et al.* 2002). Increasingly, analyses of large SNP data sets, such as the recent analysis of 2,576,903 human SNPs (Zhao and Boerwinkle 2002), are proving to be valuable for studies of mutation bias. The availability of SNP data from many different taxa now makes it feasible to develop a more detailed knowledge of factors that contribute to variation in mutational biases.

A number of analyses of mutations have demonstrated that context, or the composition of nucleotides flanking a mutation, can have a significant influence on both mutation bias and overall mutation rate (Bulmer 1986; Morton 1995; Krawczak *et al.* 1998; Zhao and Boerwinkle 2002; Morton 2003). Although context effects are not often considered in studies that apply mutation parameters (although see Arndt *et al.* 2003; Siepel and Haussler 2003), there is evidence that understanding and incorporating such effects may be very important for interpreting genomic data (Morton 2003; Siepel and Haussler 2003) since they can result in variation in mutation dynamics across sites. In nuclear genes, the most apparent neighboring nucleotide effect that has been studied to date is the CpG effect, which is an increased rate of transitions at CpG dinucleotides as a result (145) of deamination of methylated CpG sites (Duncan and Miller 1980; Bulmer 1986; Cooper and Youssoufian 1988). The CpG effect has been primarily studied in vertebrate genomes (Krawczak *et al.* 1998; Zhao and Boerwinkle 2002; Fryxell and Moon 2005), and in human sequences there is a fivefold increase in the rate of transitions at CpG sites due to deamination of methylated cytosines (Krawczak *et al.* 1998).

The CpG effect appears to be weaker in G + C-rich regions, possibly due to greater local helix stability (Fryxell and Moon 2005), and appears to be slightly stronger on the coding strand than on the template strand near genes (Krawczak *et al.* 1998).

Context dependency of mutations has also been studied in grass chloroplast DNA (cpDNA; Morton 1995, 2003). In this genome there is a significant correlation between the A + T content of the two sites flanking a mutation (the A/T context) and both the overall substitution rate and the transition: transversion (Ts:Tv) bias, due to a decreasing rate of transition substitutions as the A/T context increases (Morton 2003). Since the observed context dependency is not consistent with CpG deamination, and since CpG methylation is not known to occur in cpDNA, it has been suggested that factors such as polymerase fidelity and variable repair efficiency may be responsible for context-dependent mutation biases (Morton 2003). Neighboring base composition also influences substitution dynamics in cpDNA in other ways; both the bias toward A + T and the bias toward pyrimidines are a function of context (Morton 2003). Similar context-dependent mutation patterns appear to exist in cpDNA across different flowering-plant lineages (Morton 1997; Yang *et al.* 2002).

Given the growing body of evidence regarding context dependency and the lack of data about regional variation in mutation properties, there is a need to better understand context dependency and how mutation dynamics vary across individual genomes. To further our understanding of mutational context and variation, we have analyzed a large SNP data set generated from nuclear genes of maize (*Zea mays* ssp. *mays*) with respect to both regional and flanking base composition. We find evidence that the A + T content of flanking nucleotides has an influence on various aspects of mutation dynamics and report a correlation between regional base composition and both CpG effect and the relative rates of GC→AT and AT→GC mutations, or GC→AT mutation pressure.

Discussion

The SNP analyses presented here yield some of the first data about context and variation in mutation dynamics within a genome. They demonstrate that context has a significant influence on mutation dynamics in maize nuclear DNA: there is a relationship between flanking base composition and mutation bias, an increased rate of transitions at CpG dinucleotides, and a relationship between regional base composition and GC→AT pressure. We should note that a number of our observations are based on polarizing mutations. For our analyses we polarized mutations by using the majority base at each site to infer the original state. This will not affect the analyses concerning flanking base effect on rate and transition bias and, therefore, the overall conclusions about context effects. In addition, although the polarization does allow us to infer the mutation rate away from CpG dinucleotides and provides stronger evidence, the high rate of transitions at these sites is in itself strong support for a CpG effect. Conclusions based on predicted equilibrium composition and GC→AT pressure are, however, fully dependent on polarizing the mutations that allow us to generate the 4 x 4 matrices. Future analysis using an outgroup taxon will allow us to examine these effects and to assess the validity of using the majority base to polarize mutations.

The most notable context effect is an elevated rate of CG→TG and CG→CA transitions relative to other transitions. Given the existence of CpG methylation in plants, this rate elevation is most likely the result of a deamination of methylated cytosines at these dinucleotides. It is difficult to compare the magnitude of the CpG effect observed here directly to studies of nonplant taxa since methodologies differ, but it appears that the increase in transition rate that we observed at CpG sites, roughly a 2.1-fold increase relative to the transition rate at other sites, is not as high as what has been observed in vertebrates. Although we observe an overall 2.1-fold increase in transition rate due to CpG deamination, this increase ranges from a 1.7-fold increase in regions with lower A + T content (<48%) to a 2.6-fold increase in regions with higher A + T content (>60%) and shows a general increase with increasing regional A + T content.

This trend may reflect variation in the degree of CpG methylation across loci or that repair of deamination products is more efficient in G + C-rich regions. Along with a significant CpG effect, there are other influences of context on mutations apparent in our data. In particular, the composition of the two immediate neighbors, one 5' and one 3', of the mutation site is correlated with overall rate, transition bias, and GC→AT pressure.

These effects are similar to what is observed in grass cpDNA and it is likely that they are due to an influence of local composition on polymerase misincorporation or mismatch repair. The similar relationship between context and mutation properties in both nuclear and cpDNA is interesting since it suggests shared replication and/or repair processes or that these properties are fundamental to mutations. Much remains to be learned about replication and repair in plants, but it is known that the two genomes do not share the same replication machinery and have significant differences in repair dynamics. As more is uncovered about the replication and repair processes in the two genomes, we should be able to better understand the causes of similar context effects.

Although we found a correlation between the composition of the two immediate neighbors and mutation properties, we did not see a clear relationship between mutation and the composition of individual neighboring nucleotides that do not flank the mutation. This contrasts with a recent study of human SNPs. Again, however, differences in methodology make it difficult to draw any specific conclusions about differences in context effects. In our study we controlled for the composition of the immediate neighbors, something that was not done in the study of human SNPs. Thus, it is possible that the human SNP study confounded immediate flanking base effects and nonrandom dinucleotide composition.

Despite the lack of correlation between specific individual nucleotides beyond the immediate neighbors and mutation dynamics, we do observe a correlation between regional composition and GC→AT mutation pressure. It is possible that this correlation is not a context effect but a secondary effect arising from a relationship between chromosome location and replication/mutation dynamics.

Since, like our data, their observation was for non coding sequences near genes on the coding strand and is found across numerous loci, they proposed that the skew was due to a transcription-coupled mismatch repair system. If this is the case, then the similar finding in our data suggests a similar mechanism in plant nuclear genes. It also raises the possibility that the G over C and T over A skew observed along the leading strand in prokaryotic genomes is at least partially the result of a transcription-coupled repair mechanism. The possibility of a transcription-coupled repair mechanism has significant implications for our understanding of compositional bias in genes, such as codon usage bias.

10.2 Artículos periodísticos

10.2.1 First complete cancer genome sequenced

Scientists decipher each of the 3 billion DNA bases from the genome of an acute myeloid leukemia tumor.

For the first time, a complete cancer genome, and incidentally a complete female genome, has been decoded, scientists report online Nov. 5 in 'Nature'. In a study made possible by faster, cheaper and more sensitive methods for sequencing DNA, the researchers pinpoint eight new genes that may cause a cell to turn cancerous.

"Since cancer is a disease of the genome, this newfound ability to determine the complete DNA sequence of a cancer cell is enormously powerful," comments Francis Collins, a geneticist and former director of the National Human Genome Research Institute in Bethesda, Md., a group that raced to sequence the first entire human genome.

"We need to know the genetic rules of cancer," says coauthor Timothy Ley of Washington University in St. Louis. Ley and colleagues read each of the 3 billion building blocks of DNA from tumor cells in a woman with acute myeloid leukemia, or AML, a highly malignant form of blood and bone marrow cancer. Then the team compared the long string of code with one taken from noncancerous skin cells from the same woman.

This new sequencing technology, called massively parallel sequencing, makes it possible to compare the normal DNA sequence to the cancerous DNA sequence in the same patient. That, in turn, allows researchers to find individual DNA bases -the needles in a haystack of 3 billion pieces of straw- that had mutated in the cancerous cells.

Kevin Shannon, director of the Medical Scientist Training Program at the University of California, San Francisco, studies the genes that may lead to leukemia and calls this work "a major achievement," one that is "remarkable for its rigor and precision".

None of the researchers knew what to expect for the number of mutated genes in the cancerous cells. “We were flying blind,” says Ley. But after rigorously pruning the data to keep only the most significant mutations, the researchers identified 10 mutations, eight of which were in genes never before implicated in AML. Of these eight new mutations, none were found to be mutated in tumors from other, smaller-scale studies, suggesting that individual AML cases are distinct.

It may be that the disease is so specific doctors will need to sequence each individual with AML to determine the best course of treatment, says coauthor Elaine Mardis, also of Washington University.

At the same time, because those earlier studies did not sequence the entire genome, and because this new study had a sample size of only one patient, it is too early to tell if AML has different kinds of mutations in different patients.

So, equally possible is that common mutations in similar groups of genes may contribute to AML. Discovery of these gene networks could allow doctors to use these common pathways of disease to treat patients similarly.

“It’s fun to speculate,” Maris says, “but we just don’t know.”

Understanding the genetic basis of cancer could lead to highly personalized treatments, says Mardis. “Right now, they’re all treated the same way they were 25 years ago,” she says of AML patients.

It would be nice, Mardis says, if doctors could tell their patients, “Here’s what we know about your disease, and here are your best treatment options.”

Although scientists read every base pair in the patient’s genome, they only analyzed mutations in the DNA sequences that produce proteins, an estimated meager 1 to 2 percent of the human genome. To find mutations in other regions called intergenic DNA will require intensive statistical analyses. “We haven’t finished the job,” says Ley.

Because this study was designed to find genes that were mutated in a cancer genome, researchers omitted the DNA sequences from the sex chromosomes, the Xs and Ys, when making comparisons. Little is known about the differences between a male and a female genome.

The research team currently has funding to support more cancer genome sequences in the next few years. “What we need are thousands of genomes from each cancer,” says Ley. “We’ve already started a second patient, and are nearly finished, but our hopes are to do more.”

10.2.2 Macaque genome deciphered; may herald medical breakthroughs

Scientists have finished sequencing the genome of the rhesus macaque monkey in work they say will enhance medical research in a wide range of areas, including HIV and neuroscience. The findings will also advance scientists' understanding of primate evolution and what makes humans genetically distinct. (Read a genetics overview). An analysis comparing the macaque genome to the already sequenced chimpanzee and human genomes shows that the three primate species share about 93 percent of their DNA. But they have some significant differences among their genes.

"We really want to know what makes us humans -and different from our primate cousins," said Richard Gibbs, director of Baylor College of Medicine's Human Genome Sequencing Center in Houston, Texas.

"This study allows us to observe what has been added or deleted in each of these three primate genomes during their evolution." Gibbs is the project leader for the Rhesus Macaque Genome Sequencing and Analysis Consortium, an international team of more than 170 scientists from 35 institutions whose findings are published in tomorrow's issue of the journal 'Science'.

Evolutionary pressures

The rhesus macaque is the second nonhuman primate, after the chimpanzee, to have its genome sequenced. Chimps are believed to be the primates most closely related to humans. Compared to the human genome, the chimp genome is only about 1.5 percent different, while the macaque genome is about 7 percent different. That divergence makes macaques ideal for the evolutionary study of primates, because important features that have been conserved in primates over time can be more easily seen by comparing rhesus monkeys to humans than by comparing chimps to humans. "In some ways the chimp, which has already been sequenced, is a little too close for us to make easy sense of what the differences are" between humans and other primates, Gibbs said. It's a little easier to contrast the macaque with either chimp or human to make sense of what's going on," he added. "If the chimp and the macaque share a feature and the human is different, you can say this is a human change."

Researchers say that humans and macaques had a common ancestor about 25 million years ago. About six million years ago, chimps split off from the human lineage. “Once they diverge, other changes occur that reflect what the evolutionary pressures are on that particular species," Gibbs said. “What we see now are snapshots of the molecular fossils that reflect what had happened to that species since the divergence.” The researchers identified nearly 200 genes that probably play a key part in determining differences among primate species, including genes involved in hair formation, immune response, membrane-protein generation, and sperm-egg fusion.

Scientists were surprised to find some instances where the normal form of the macaque protein looks like a diseased human protein. One such example is phenylketonuria, a genetic disorder that can lead to brain damage and mental retardation because sufferers lack an important metabolic enzyme. “The underlying question here is, what is what apparently looks like a human mutation that causes a devastating disease with early demise and severe mental retardation doing in an apparently normal macaque?” Gibbs asked.

AIDS research

Because the rhesus macaque is both abundant and genetically and physiologically similar to humans, it is widely used in medical research, particularly in vaccine research and as a model for AIDS research. Scientists expect the rhesus macaque genome sequence to enhance research in neuroscience, behavioral biology, reproductive physiology, endocrinology, and cardiovascular studies. (Related: “Dog Genome Mapped, Shows Similarities to Humans” [December 7, 2005].)

Some interesting findings are already coming from a study of variations within the macaque genome. The complete genome sequencing of the macaque was done with the DNA of a single individual -a female rhesus macaque at the Southwest Foundation for Biomedical Research in San Antonio, Texas. But researchers also sequenced parts of the genomes of 16 other macaques, 8 from China and 8 from India. The analysis suggests that the two populations, from India and China, separated about 162,000 years ago.

"We're able to say that the two subspecies of the macaques are very different from each other on a genetic level, probably much more different than human populations are from each other," said study leader Carlos Bustamante, an assistant professor of biological statistics and computational biology at Cornell University in Ithaca, New York. The simian immunodeficiency virus (SIV) is used as a model for the human immunodeficiency virus (HIV). When exposed to SIV, Chinese macaques develop AIDS-like symptoms more slowly than Indian macaques, the researchers found.

"Indian animals came down with AIDS much faster than Chinese animals, so there's a huge interest in trying to figure out what are the genetic differences that may account for that," Bustamante said.

Researchers looking for a disease-causing gene don't usually find the exact location of the gene right away. Instead, the scientists first determine that the gene is situated somewhere along a particular DNA strand between two easily recognized sequences called markers, and then zero in from there.

"We found that for Indian macaques the number of markers you need to map a genetic disease will be much smaller than the number of markers that you need in Chinese macaques or even in humans," Bustamante said. "This means the search for disease-causing genes may be easier in Indian macaques." Gibbs, the overall project leader, points out that sequencing the macaque genome doesn't mean more macaques should be used in laboratory research.

"I think we're doing something quite opposite," he said.

"We're entering into a new area where we can actually do much more rational and more informed experiments with macaques. We're knowing the macaque better instead of just advocating experimenting on the macaque."

10.2.3 Researchers focus on differences between groups to find bad DNA

More than 15 million elderly Americans slowly lose their eyesight due to age-related macular degeneration: an accumulation of inflammation-related protein and fat beneath the center of the retina that slowly destroys it. The disorder runs in families, but the gene responsible had eluded scientists. In March three separate teams announced that they had zeroed in on a DNA sequence on chromosome 1 that carries the gene for complement factor H, a protein involved in regulating inflammation. A mutation in this gene may account for about half the cases of macular degeneration in the United States.

This genetic culprit was revealed by a second and little-heralded phase in the Human Genome Project and it comes five years after a rough draft of the entire human genome was announced. Touted as the key to deciphering the genetic book of life, that initial sequence has proved most useful for finding or confirming genetic mutations that cause rare diseases such as Tay-Sachs disease and Huntington's. These alterations are relatively easy to identify because they can be traced and isolated in families with a history of the disease. Finding genetic clues to common diseases is much more difficult because many genes -as well as lifestyle and environmental exposures- may be involved. So rather than search the entire genome for genes related to common cancers, heart disease, asthma, and diabetes, scientists have turned to detecting inherited variations in the genomes of different populations and how they may be linked to disease vulnerability.

The human genome contains 6 billion nucleotides, but the differences between one person's DNA and another's are slight. Still, the scattered differences that are common to various populations provide a map for hunting down disease. Looking at whether shared variations -single-letter misspellings- in vast chunks of DNA called haplotypes are correlated with certain diseases can help researchers decide which segments of the genome to scour for more specific clues. The project, called the HapMap, is the effort of a six nation consortium. The group has spent more than two years analyzing DNA from 269 donors with diverse ancestry. They announced their first major milestone -identifying 1 million of these common variations- in February.

To find the macular degeneration gene, researchers compared DNA from people with and without the disease. All three teams found sections of DNA -haplotypes- that differed and ultimately pinpointed a single-letter difference that changed the amino-acid content of complement factor H. HapMap researchers say that refining the map further will speed up such discoveries, and they plan to release a new version this month that will include 4 million single-nucleotide variants.

For now, the HapMap project does not mean new genetic tests for diseases are about to appear, but at least one company is considering whether to develop a test for genetic vulnerability to macular degeneration.

Although having the mutation more than doubles a person's risk of the disorder, which causes loss of central vision, it doesn't mean that a person will automatically develop the disorder. Moreover, because no one is certain about how to minimize risk, a test wouldn't have much practical value, according to Stephen Daiger, a geneticist at the University of Texas Health Sciences Center.

Yet there is good news for some patients. In July Genentech announced that the drug Lucentis stabilized or even improved vision in a yearlong clinical trial of patients with so-called wet macular degeneration, a form that involves overgrowth of blood vessels. The drug works by inhibiting a protein involved in blood-vessel formation. Lucentis is one of a handful of other drugs that combat this form. No drugs are effective for the other, far more common form of the disease.

Alzheimer's: of the three variants of the ApoE gene, ApoE4 carries the highest risk for late-onset Alzheimer's (other genes are associated with early-onset Alzheimer's). No cure or effective treatment exists for Alzheimer's, so ethicists usually advise against genetic testing unless there is a strong family history or a diagnosis is in question.

Breast cancer: out of more than 200,000 new cases of breast cancer each year, about 10,000 or more are linked to the BRCA1 or BRCA2 gene. Testing positive for the gene means you have at least a 35 percent chance of developing breast cancer. People with a family history of the disease use the test results to help them decide what steps to take to preserve their health.

Colon cancer: about 7,000 patients a year have a hereditary colon cancer associated with one or more gene variants. All of these variants are linked to the onset of colon cancer by age 45 in most carriers. Testing may be advisable for people with a strong family history of colon cancer in order to begin screening and management.

Although tests for many rare genetic disorders such as Tay-Sachs disease and Huntington's disease are available, only a few tests can predict the risk of common diseases. And even where studies have linked a gene to a disease, people carrying the gene or genes will not necessarily develop the disease; they simply face an increased risk. Nor does not having the gene mean the person will remain disease free, because many diseases result from interactions of multiple genes and the environment. Below are examples of genetic tests that are currently available and how predictive each is for individuals. Genetic counseling should accompany these tests.

Dialogue
Genes are not destiny

David Altshuler, director of medical and population genetics at the Broad Institute of Harvard University and MIT and an associate professor of genetics and genetic medicine at Harvard, is hunting for genetic clues to diabetes and prostate cancer.

Can population genetics -and genetic testing of individuals- predict if someone will get a disease?

A: Population genetics is the study of variations in heredity that run in families. We try to find a gene or genes responsible and learn how they cause the variation. It is seldom predictive. Having a gene associated with a disease is an indicator only that you might get that disease. It's like taking a cholesterol test. It's a warning that the odds are higher than normal. It can indicate a change in diet or lifestyle or an intervention with a drug.

How accurate is the information?

A: Genetic markers are found all the time that come from comparing people with a disease to those without the disease, but until you figure out what the gene does, it is not useful information.

There are a lot of hucksters out there saying that this or that gene will predict a disease- it's not true for most common diseases. You have genes that definitely cause rare genetic disorders like Huntington's. And tests for genes for diseases like breast cancer have been validated. But for most common diseases, the data need to be thoroughly understood and the links validated by having many labs replicate the findings. Many times the findings are relevant only to a particular population where the test was done. People should be skeptical of most of these tests.

Does sequencing the genes of animals such as the mouse and the dog help us understand ourselves?

A: If you want to understand human genetics, you need to line up a lot of species to compare genes to find how similar and different they are. This tells us how these genes evolved and gives insight on what the genes do and how mutations might be treated with drugs.

Will we ever have a little card we carry around that has our genome on it?

A: We could already carry cards like that with family histories or diseases and our latest blood tests, but we don't. I'm less interested in tests than treatments.

10.2.4 The 1 percent genome solution

Tiny slice of genome reveals bustling activity in the gaps between genes.

The first results from a massive project to exhaustively catalogue all the functions of the human genome reveal a hotbed of activity in the gaps between genes. An international consortium of researchers sifted through 1 percent of the genome looking for pieces of DNA that are copied by the cell or help to control gene activity. The results indicate that most DNA is copied into molecules of RNA, including the long stretches between genes, and that genes overlap and interact with each other much more than researchers previously believed.

“We all suspected there was interesting stuff going on in these regions [between genes], and sure enough there is,” says bioinformatician Ewan Birney of the European Bioinformatics Institute near Cambridge, England, a member of the project’s computer analysis team.

Although researchers do not yet know the biological significance of these discoveries, they say that fully cataloguing the genome may help them understand how genetic variations affect the risk of contracting diseases such as cancer as well as how humans grow from a single-celled embryo into an adult. The next phase of the project, set to begin later this year, will attempt to inventory the full genome.

A genome consists of only four different nucleotide bases, or DNA subunits, arranged in a particular sequence. The publication of the human genome in 2001 revealed its sequence -the significance of which remains a mystery. In particular, genes account for only 1.2 percent of the genome's three billion bases. Once dismissed as “junk DNA”, researchers have found that some of these so-called noncoding regions are shared among mammals, suggesting they play an important function. To help uncover those functions and identify other important sequences, 35 research groups joined forces in 2003 to create the encyclopedia of DNA elements (ENCODE) project. This consortium selected 44 separate sections of the genome that included regions of high to low gene density and high to low similarity between mouse and human.

Like treasure hunters combing a vast beach with metal detectors, ENCODE researchers sifted through their patch of the genome in multiple ways that are described, along with the results, in a ‘Nature’ paper published online today and in a special issue of ‘Genome Research’.

A major part of the project was identifying sequences that cells copy, or transcribe, into RNA molecules. Cells make proteins from RNA they copy from genes, but some RNAs play roles by themselves. In addition, some studies have found evidence that species from flies and worms to humans copy large amounts of RNA from noncoding DNA, with no apparent purpose. Nevertheless, “before ENCODE, I think a lot of people were skeptical of how real intergenic activity was,” says bioinformatician and consortium member Mark Gerstein of Yale University.

Although genes make up only 3 percent of the ENCODE sequence, the consortium found that 93 percent of the sequence is transcribed. Some of the transcripts hail from noncoding DNA, the researchers report, but those that do match up with the 399 ENCODE genes overlap with each other extensively.

Transcripts from 65 percent of the genes incorporate pieces of DNA from relatively far outside of the genes or even from one or two other genes, says molecular biologist and consortium member Tom Gingeras of Affymetrix, a genome technology company in Santa Clara, Calif.

Researchers know that cells chop single genes into shorter pieces called exons, which they mix and match into one transcript for creating a protein. Gingeras says the ENCODE findings confirm recent reports that humans and flies sometimes combine exons from two different genes. Based on the transcript sequences, the researchers identified 1,437 new promoters -short DNA sequences where transcription begins- in or between genes, on top of the 1,730 promoters they knew of. That is nearly ten promoters per gene, Birney says. He adds that the abundance of transcripts that overlap each gene suggests that the very term ‘gene’ should mean something different inside the cell nucleus, where transcription takes place, than outside of it, where finished proteins go.

Project members also catalogued sequences that mark areas where DNA unwinds from the round histone proteins that maintain the shape of chromosomes, allowing the cell's transcription machinery to activate genes in those areas. They discovered some potentially unwound areas that are far from promoters and may therefore play some other role, Birney says.

The consortium found that 5 percent of the studied sequence has been conserved among 23 mammals, suggesting that it plays an important enough role for evolution to preserve while species have evolved. But of all the new ENCODE sequences identified as potentially important, only half fall into the conserved group.

These unconserved sequences may be "bystanders," Birney says -consequences of the genome's other functions- that neither help nor hurt cells and may have provided fodder for past evolution.

They could also simply maintain a useful DNA structure or spacing between pieces of DNA regardless of their particular sequence, says genomics researcher T. Ryan Gregory of the University of Guelph in Ontario, who was not part of the consortium.

"The biological insights are mainly incremental at this point," says genome biologist George Weinstock of the Baylor College of Medicine in Houston, which he says is to be expected of such a pilot study. "This is a 'community resource' project, like genome projects, that makes lots of new data available to the community, who then dig into it and mine it for discoveries."

Gregory says the results, although still cryptic, do hint at new functions and a more complicated genome. "This study shows us how far we are from a comprehensive understanding of the human genome."

10.2.5 The next frontier: proteomics

Even before they finished decoding the human genome, scientists began the next and far more challenging step in explaining the molecular underpinnings of life. It’s called proteomics -the cataloging and analysis of every protein in the human body.

Although proteins are the direct result of the instructions coded in our DNA, they are far more variegated and complex than DNA. They have to be. Every chemical reaction essential to life depends in one way or another on their services. Proteins are the beams and rafters of the cell and the glue that binds the body together; they're the hormones that course through our veins and the guided missiles that target infections; they're the enzymes that build up and break down our energy reserves and the circuits that power movement and thought.

What a protein does is largely determined by its shape. Proteins are stippled with pockets and grooves into which molecules fit as precisely as a key fits a lock. To fully understand how a protein works, you have to be familiar with every nook and cranny on its surface, which is why the National Institute of General Medical Sciences will spend $20 million this fall to establish a series of research centers dedicated to a branch of proteomics known as structural genomics. The centers will detail, over the next 10 years, the shapes of 10,000 proteins. That’s a tiny fraction of all the proteins found in nature, but the NIGMS thinks that number will cover most of the structures relevant to biology and medicine.

Why not study all the proteins? For one thing, there are too many -50,000 to 2 million, depending how you count. For another, most of those millions of proteins are just variations on a handful of themes. Proteins with similar functions -be they in insect, worm or man -often share structural characteristics that are reflected in the genes that encode them. Structural biologist Stephen Burley of Rockefeller University estimates that the maximum number of distinct shapes may be as few as 5,000.

The NIGMS hopes to construct a lexicon of shapes -barrels, doughnuts, globular spheres, molecular zippers and so on -that when mixed and matched will spell the shape of any gene's product. About 1,000 of these structures -and the genes that code for them- have already been cataloged.

But naming the shapes and knowing precisely how they are formed are two different things. Proteins twist and pleat themselves as they're synthesized. These basic forms are then further folded and linked to other proteins to create the superstructures crucial to protein chemistry. Scientists traditionally dissect the atomic details of these folds by observing how crystallized proteins scatter X rays- experiments that can take years to complete. But robots and powerful X-ray generators have lately boosted the pace of discovery. Structures that two decades ago would have taken a couple of researchers 10 years to crack can now be solved by one in a matter of weeks. "By the end of the five-year pilot phase," predicts John Norvell, director of the NIGMS program, "each of the centers will be producing 100 to 200 protein structures a year."

That may not be fast enough. Already, a host of biotech firms -including PE Corp., the parent company of Craig Vent's Celera- have launched their own proteomics programs, some focused on protein structures, others trying to determine where in the body different proteins are produced and how each is controlled. The flurry of private activity raises the specter of intellectual-property disputes like those that plagued the Human Genome Project.

Last spring the NIGMS co-sponsored the first international structural-genomics meeting, partly to nip those conflicts in the bud. The hope is that history will repeat itself a little differently this time around.

10.3 Editoriales

10.3.1 DNA: one teacher's reflection

This issue of 'Science' appears amid a swirl of contemporary reminiscence about DNA, triggered by the 50th anniversary of James Watson and Francis Crick's paper in which its structure was proposed. What they gave us was a model that derives its beauty from the way in which it fit with what was already known about the biology of genetic control and the chemistry of DNA's nucleotide components.

Of course, there was a splendid storehouse of information already available to provide that resonance. Sometimes we forget how much; I now encounter undergraduates who believe that Watson and Crick 'discovered' DNA -a claim those authors would be quick to disavow. In fact, much was known about the biochemistry of DNA; in particular, Chargaff's demonstration that across species, A = T and G = C. The experiments of Avery, McCarty, and MacLeod in 1944 in bacteria, followed by those of Hershey and Chase in bacteriophage, had already demonstrated that it was DNA and not protein or something else that was the molecule of heredity. These experiments provided the background behind Watson and Crick's interpretation of the x-ray crystallographic data of Rosalind Franklin. The Watson-Crick model appeared as I was becoming a novice teacher of introductory biology. Later, as a physiologist trying to learn and then teach the 'new biology', the unfolding story struck me as a kind of marvel. Meselson and Stahl fed bacteria with nucleotides labeled with heavy nitrogen and then used density-gradient centrifugation in cesium chloride to demonstrate the semiconservative nature of replication. Kornberg showed that DNA polymerase could support replication in a cell-free system if only it was given some 'primer' DNA and a supply of the four nucleoside phosphates. From that point on, the transcription part of the problem shifted to the molecular mechanisms of unwinding and replication, to the knotty problem of how damage is repaired, and to the regulation of transcription.

Translation became the key issue as the 'central dogma' established the sequence of complementary synthesis that ran from gene to messenger RNA to protein, and led to the deciphering of the code through working out the associations of transfer RNAs and specific amino acids. That, too, depended on some much earlier biology that I came to know through historical proximity. In a Stanford basement laboratory, Beadle and Tatum switched abruptly from 'Drosophila' to 'Neurospora' and soon developed screens for nutritional mutants to show that single mutations blocked specific enzymatic steps in a synthetic pathway. Later work continued to be informed by the 'one gene, one enzyme' concept, in a worldwide explosion of results that exposed the roles of various RNA species in the task of protein synthesis.

Making maps (that is, establishing the relation between DNA sequences and the location of genes) began in microorganisms through work on bacteria by Lederberg and Tatum and on viruses by Benzer's group. To do the same in eukaryotic cells proved a more difficult challenge. Many years later, I could hang around and watch as Yanofsky's group used the same organism as Beadle and Tatum had used in the same basement to establish the co-linearity of the amino acid sequence of an enzyme with the nucleotide sequence of the gene encoding it.

A biochemistry teacher of mine was fond of pointing out that if we look deeply enough into the cell, chemistry becomes anatomy. The Watson-Crick structure achieved that fusion for the front end of the central dogma, by uncovering the spatial geometry that makes replication and the conservation of information work in the cell. At the other end, we have now learned from the Noller and Steitz groups that the spatial geometry of the ribosome makes possible the final assembly required by the instructions originally contained in the Watson-Crick structure. In an interesting surprise, the enzyme responsible for this last task in translation is not a protein but RNA itself –an odd exception to the central dogma. Surely that tells us that the book is not yet closed and that the line of work begun 50 years ago will continue to yield the unexpected. For example, who would have expected a decade ago that a brand-new family of small RNA molecules would be able to control gene expression? Will that be the anniversary we celebrate in 2053?

10.3.2 Human genomes, public and private

The burgeoning commercial sector that is based on genome information poses a challenge to the norms of scientific publication. But it remains to be established that the conditions of access to published sequence data need to change.

The human genome sequence contains the genetic code that sits at the core of every one of the ten trillion cells in each human being. It profoundly influences our bodies, our behaviour and our minds; it will help the study of non-genetic influences on human development; it will unlock new insights into our origins and history as a species; and it points to new ways of combating disease. The people of many countries have invested in the Human Genome Project's determination of the sequence, and it is hard to see how that investment could have received better returns. Having released their data daily from the outset with unrestricted access, the publicly funded consortium has assembled about 92% of the sequence. 'Nature' is delighted this week to publish the project's analysis, and related results, freely available to all without restriction at http://www.nature.com and on pages 814–958.

In so doing, 'Nature' has followed a traditional model in the publishing of extensive scientific data. As indicated in our 'Guide to authors', we require the results of genome sequence analyses, as with protein structure coordinates, to be immediately available from an appropriate database without restriction. This supports an unwritten contract with our readers that what they see described is what they can use, without obstacles, whether they work in the commercial or academic sector (an increasingly blurred distinction). It supports a broader principle by which scientific results are available for searching and use with software tools. And it supports a principle enunciated by the United Nations that the human genome in particular is, in a symbolic sense, humanity's common heritage.

It is worth noting that 'Nature' sees a distinction between access to essential data embodied in a paper, and access to the materials and techniques used. Making freely available all the materials used for a piece of research is sometimes impractical. All we can do is to sustain a policy of access to materials as far as possible, and to cooperate in attempts to set up new standards of access as technologies develop.

Money makers

What, then, of companies that make a living from data discovered through their own research? There is big money at stake. The bioinformatics market alone will soon exceed $1 billion per year. Some pharmaceutical companies are valued at hundreds of billions of dollars, much of which depends on their ability to exploit genomic information in developing blockbuster drugs. According to stock market commentators, there are more than 100 drug-platform biotech companies chasing some $3 billion of research and development money.

Genomics-based drug companies such as Human Genome Sciences, Millennium and CuraGen are forming partnerships with big drug companies to help them grow. Companies such as Affymetrix, Celera, Gene Logic and Incyte are spending billions on software and hardware. Information-technology giants such as Motorola and IBM are increasingly bringing their expertise to bear in the supply and analysis of microarrays and other biochips.

How do information generators and curating companies make their living? Celera, a company in the vanguard of those seeking to use the human genome sequence to make money for private investors, depends on much more than the sequence to pay its way. Its sources of revenue include subscriptions, software, third-party licence revenue, consulting and diagnostics. Its total revenues are estimated at $130 million, and its capitalization stands at over $1 billion. Some analysts are concerned about Celera's future profitability in the face of the availability of free data from public projects. Others are confident that it will be able to add value to the basic information in a way that drug companies will pay substantial sums for. The company is gearing up to become a proteomics powerhouse and, as its president Craig Venter has said, there are too many challenges around for it to make sense for private and public scientists to duplicate one another's work.

Data access

The private sector is essential to the improvement of human health, and its investors need financial returns. Yet many of the commercial companies, whatever their motives and products, do outstanding basic science along the way.

The amount of valuable scientific information residing in research and development centres in the private sector is huge -the global R&D budget of pharmaceutical companies is estimated at $28 billion per year.

It is in the interests of all researchers that as much of that information as possible finds its way into peer-reviewed publication. Suppose a journal were handed a wonderful and unique piece of science, but that, in the interests of the originators, some of the information had to be kept behind a wall of subscriptions and licensing arrangements. In that case the conflict of interest would be acute: on the one hand, the world is better served if the information is available, even with conditions, than if it is not available at all; on the other hand, such a situation not only breaches the traditional norms of scientific publication, but also hampers the ability of science to progress rapidly by the unrestricted reanalysis of published data.

Since we established our policy on access to genome data in January 1996, 'Nature' has been able to hold the traditional line. The burden of providing proof that the line should be abandoned lies with the companies -either through rational debate or possibly by the sheer scientific significance of their output in the absence of a publicly funded equivalent. With a publicly funded project delivering data, 'Nature' believes that the human genome sequence is not the place for the traditional rules to be broken. We are willing to cooperate with competing journals to maintain the model of open access as far as possible, as we have in the past. We are also willing to engage in consultation where new types of privately generated data, lacking their equivalent in the public domain or in standard public databases, require new thinking. But, in the meantime, the more that both private and public activity can stimulate each other with freely available science, the better it will be for everyone.

10.3.3 Our human genome -how can it serve us well?

Good public health, like all good medicine, rests on empirical data of high quality. If we have records of the incidence of disease, and of how this responds to interventions, we will be able to learn the best approaches for each community. Nowhere is this more true than for genetic diseases, where medical knowledge is increasing rapidly and personal issues, cultural differences and ethical dilemmas abound.

Until recently, it was possible to dismiss genetics as an issue affecting a few children in rich countries. Genetics looked unimportant compared to infectious diseases, high infant mortality and lack of proper sanitation. While those problems have not gone away, there are many countries (including India and China, the two largest in population) in which most people face the same medical issues as in ‘first world’ countries -cancer, heart disease, psychiatric disorders, diabetes, and Alzheimer’s disease.

With this shift, the burden of genetic disease has changed dramatically. The adult diseases listed above all have a large genetic component, often due to changes in more than one gene, which interacts with the environment in susceptible individuals.

Even the consequences of cigarette smoking, one of the most serious public health issues in developing countries, are not uniform but vary with genetic susceptibility. Greater knowledge of genetic factors that lead to high risk, which must be acquired for each ethnic group separately, will help public health planning and the targeting of interventions where they are most needed.

Among children, once a country has lowered mortality and morbidity due to poverty, genetic diseases due to single gene mutations such as haemoglobinopathies, cystic fibrosis and muscular dystrophy consume a large proportion of pediatric resources. Of these diseases, the haemoglobinopathies (sickle cell anaemia and thalassaemia) are by far the most important internationally.

There are about 200 million healthy carriers of thalassaemia or sickle cell anaemia. These carriers have a small added resistance to malaria in infancy, which is why the mutations causing haemoglobinopathies have spread throughout countries in tropical and sub-tropical regions where malaria is common. If two carriers have a child, it has a one in four chance of inheriting the mutation from both parents, causing a serious anaemia. Over 300 000 children each year are born with a severe haemoglobinopathy. With worldwide migration, these diseases are as much a feature of Europe, the United States and Australia as of the countries where they originated.

In this issue, Professor Bernadette Modell, who pioneered the reorientation of clinical genetics to public health and coined the term 'community genetics', puts forward evidence on the usefulness of genetic registers, in this case for thalassaemia in the United Kingdom (pp. 1006-1013). It is fascinating that even in the UK, where there is a highly organized national health system, the quality of the data depends upon the curator of the genetic register ('Personal contact is the key to success').

It is a pity that Professor Modell did not comment on this point, since it will be important when judging transferability to a country where transport may be poor and the phones may not work well. However, it is encouraging that in the UK there were no problems about confidentiality, and families were confident that the register would be used in a positive and not in a discriminatory way.

What of the ethical dimension? WHO is at last beginning to take a lead in offering policy on ethics of 'the new genetics', as can be seen from consultation statements on 'Genomics and world health' and on 'Ethical, legal and social implications of genomics' issued during 2001 by WHO's Advisory Committee on Health Research. How do these statements fit with Professor Modell's argument for better data registers?

Some of the most sensitive issues in genetics relate to the ethics of pregnancy choice. Some believe it is unethical for a country to push families towards screening, with the implication that an abortion may be a choice. Others believe it is equally unethical to deprive a couple of that information and force them to have a child who will be seriously handicapped with a condition for which there may be no effective therapy.

Choices by individual families can only be made if they are legal and the resources are there to provide the information needed. In the UK, where prenatal diagnosis is available, some ethnic groups used this option, while others did not. This would be completely appropriate provided that the choices were made by the couples themselves in the context of their beliefs and in good time, but unfortunately the UK study reveals that many couples who did not use prenatal testing were not offered it, which is surely a denial of their human rights to make their own decisions.

Now that the sequence of the human genome is available, it should be possible to ensure that every family has reproductive choice in relation to serious handicap. This does not in any way put down those with a handicap, any more than reducing the incidence of spina bifida by adding folate to bread reduces our determination to do our best for the smaller number of persons with a neural tube defect who are still born today.

WHO, as is appropriate, has recognized both the crucial importance, of the understanding coming from the human genome project for human health, and the issues that arise from these data for world health in general and developing countries in particular. Much of genome science is low technology and could easily be used by every country were it not for the patenting of DNA sequences, an issue on which indigenous people are particularly sensitive as ‘their genomes’ may give vital clues to susceptibility genes for common diseases such as diabetes.

Genes are about health. It is foolish to talk of cancer genes, or disease genes, or even thalassaemia genes. Most genes are healthy genes. They code for proteins and functions that allow us to survive and, usually, flourish. Their rich diversity ensures both our endless and wondrous variety as people, and our evolutionary survival. The human genome project helps us to know the power of the genome for humankind, for all our people, for now and the future. Appropriate data will allow us to harness that power better.

10.3.4 Post-genome blues

Will 26 June 2000 be remembered as one of the defining moments in the history of science? After all, this was the day that the formal announcement of the completion of the rough draft of the human genome was made. On the morning after, the newspapers were full of headlines screaming of the benefits of the project to humanity. In the days that followed we were treated to extraordinary visions of cures for every conceivable disease, whose cause lies in our genes. Among the more fanciful headlines to which we were treated was one that read -'People can live for 1200 years'; a view that appeared to be copyrighted by the Times Newspapers emanating from London, quoting the apparently unimpeachable source of one of the British governments top scientists. Another headline that labelled the genome announcement as the 'biological equivalent of the moon landing', inevitably brought back memories of the almost forgotten days of NASA's missions to the moon. Would the declaration that the human genome sequence was almost at hand, compare in impact with the moment when Neil Armstrong's voice drifted eerily across space, immortalizing the phrase -'one small step for man, one giant leap for mankind'? Would it compare with the moment when Robert Oppenheimer and his troops watched the mushroom cloud develop over the New Mexico desert, as they confirmed with a definitive experiment the destructive power of the atom? This comparison with the moon landing or the first atomic explosion might in fact serve to temper some of the mindless optimism that has found its place in the media after 26 June. In the late 1960s it did appear that a new era of rapid exploration was almost around the corner. A manned Mars landing was considered a definite possibility by the turn of the 20th century. What has happened is too well known to recount here; with [recent technical difficulties with Mars probes reminding us that 'there's many a slip between the cup and the lip'.

Unlike the moon landing or the decisive and terrible conclusion of the Manhattan project, the timing of the official announcement of the 'rough draft' of the human genome appears to be a carefully orchestrated, public relations exercise, to counter the increasingly acrimonious race to complete the genome sequence.

The competitors were a publicly funded consortium spearheaded by the US National Institutes of Health and the Sanger Centre in Britain on one side and a private company, Celera Genomics on the other. The long arm of the US presidency appears to have brokered a peace, leading to the timing of the announcement on 26 June. The cover of the 3 July issue of 'Time' features the generals, Craig Venter of Celera and Francis Collins of NIH. More famous truces, Henry Kissinger and Le Duc Tho of Vietnam fame or Anwar Sadat and Menachem Begin of the Middle East Wars, have been blessed with Nobel prizes, albeit for peace. It is not inconceivable that this will happen again, only it will be for medicine. While the sequencing of the approximately 3.2 billion bases which constitute the human genome is a major scientific achievement, a technological 'tour de force' that will be unmatched for sometime in biology, the circumstances of its public declaration of success suggest that the political and commercial implications of the project may take a long time to be fully appreciated. From a purely scientific standpoint, biology research stands to benefit enormously from the availability of the genome sequences of diverse organisms. The work of annotation and the task of establishing connections between genotype and phenotype will keep large numbers of scientists busy for many years to come. The most worrisome aspect of the genome sequencing program is that overriding commercial interests will soon ensure that the entire area of genomics and its applications will soon get mired in a sea of patent problems. The Americans have a characteristically pithy statement that summarizes the problem: 'There's no such thing as a free lunch'.

While our newspapers and TV programs have featured some of our more visible and voluble scientists talking about the broad implications of the human genome sequencing effort, few have wondered about the consequences for biological and biomedical research in this country. India has not been a partner in the international sequencing venture which has involved laboratories in France, Japan, Germany and China in addition to the major role of UK and US laboratories. How well organized are our laboratories to make use of the enormous body of sequence data that is being generated, not only human but also pathogen genomes?

Here, we usually take recourse to trumpeting the debatable claim that we have an enormous reservoir of trained computer whiz kids, who will see in genomes what no one else has seen before. The reality, of course, is that even the task of annotation is one that has been anticipated elsewhere, years ago when the sequencing programs were in their infancy. The result is that we are likely to be very far behind even in the task of building future research on the emerging genome data. A most distressing feature that has come to the fore is the facility with which we tout our large (almost impossibly so) population and ethnic diversity as a major resource in biological research in the area of genes and genomes. Even in scientific circles there is a worrying tendency to be self-congratulatory about our achievements in biological research (other areas have similar problems, although the magnitude of unsustainable claims appears to be correlated to the money spent). In order to compete effectively in the 'big science' of genomics, we need to have a hard-headed and unsentimental assessment of our strengths and weaknesses. The almost complete absence of high quality medical research that uses the techniques of modern molecular biology is a matter of serious concern. In many areas we have been spectators to technological advances that have revolutionized whole areas of human activity. It is not often recognized that our repeated drum beating on information technology (IT) rests on the fragile fabric of a non-existent hardware base. What is true of IT today will undoubtedly be true of biotechnology tomorrow.

But these depressing thoughts may be part of the 'morning after syndrome'. I should end these musings by recalling that even the big science of genomics had humble beginnings. The heart of the genome project is the procedure for reading the sequences of the letters of the DNA alphabet, the 'dideoxy procedure' developed by Frederick Sanger at Cambridge in the 1970s; later modified and automated to produce the machines that according to James Watson, could be 'run by monkeys'. Sanger himself, the recipient of two Nobel prizes (the first for sequencing proteins in 1958 and the second for sequencing nucleic acids in 1980), remained characteristically modest.

In a marvelously understated autobiographical essay (‘Sequences, sequences, and sequences’, Sanger, F., *Annu. Rev. Biochem.*, 1988, 57, 1) he says: ‘Of the three main activities involved in scientific research, thinking, talking and doing, I much prefer the last and am probably best at it. I am all right at the thinking, but not much good at the talking. ‘Doing for a scientist implies experiment.’ One would hope that there are many unsung Sangers in our midst waiting to bloom as genomics flowers in the future.

.

10.3.5 The ‘finished’ landscape

In 1990, the Human Genome Project (HGP) launched a 15-year plan to sequence the human genome. More than 2,000 scientists from over 20 institutes in six countries collaborated to produce a first ‘working draft’ of the human genome. This was a landmark in scientific research and a momentous occasion for ‘Nature’, which was the scientific journal with responsibility for the review and evaluation of the public effort, and [had the privilege of publishing the resulting paper in February 2001.

Although this might have felt like the end of an era, it was only the beginning. The draft sequence had ~150,000 gaps, and the order and orientation of many of the smaller segments had not been established. The work of the HGP laboured on, and it was 3 years later that the finished sequence was published in ‘Nature’ with 2.85 billion nucleotides and just 341 gaps.

While the HGP worked towards presenting a finished landscape of the human genome, the smaller teams that were sequencing the individual chromosomes continued their detailed analyses; the collated information from these groups would provide the most comprehensive analysis of the human genome. The first and one of the smallest chromosomes -human chromosome 22- was published in December 1999. This seminal work has since been followed by the publication of the remaining chromosomes, with the largest -human chromosome 1- being the last to be published in the spring of 2006. In keeping with the spirit of the genomics community work ethic, ‘Nature’ has made each of these papers freely available on the Internet. One could be forgiven, especially the non-genomicists, for asking why ‘Nature’ continued to publish every chromosome, especially after the ‘finished’ human genome sequence had been published. This takes me back to the autumn of 2001, when Philip Campbell, the Editor-in-Chief of ‘Nature’, Carina Dennis, the then Genetics Editor of ‘Nature’, and myself were sat in the ‘Nature’ offices discussing how to publish the individual chromosomes. ‘Nature’ had been able to publish the draft sequence of the human genome as a single event, and, while the completed chromosome analyses merited a similar single-event publication, it was clear that they could not be coordinated in the same way.

The smaller groups spanned the world, and each chromosome team was working on a unique project in terms of the level of analysis required for completion. It was important that each team had the time to complete their analysis to the highest level and to produce the best possible product without the work being lost in a competitive race to publish first, as is often the case in scientific research.

'Nature' fully appreciated the scientific value of publishing all 23 pairs of chromosomes as a single entity. Engaging in the community spirit of collaboration with which the draft sequence had been completed, 'Nature' agreed to treat the human chromosome papers as a single event, but with staged publication. The offer was for 'Nature' to review the work with the usual rigour bestowed on all of its scientific submissions, and the papers were accepted for publication only once the reviewers' queries were satisfied. The bar was raised regarding the level of analysis required, with every publication resulting in better annotation and greater biological insights. The authors went to great lengths to make each paper as engaging and accessible as possible. We have waited 6 years to publish this set of papers, which I must admit we naively envisaged would take only a year or two. Seeing the entire set of human chromosomes collected together brings home what a momentous task this has been.

The current collection starts with an overview by Francis Collins of some of the milestones of this enormous endeavour, from the inspirational agreement made by the researchers involved to place all of the data into the public domain, to the many doors that this sequence information has opened and is continuing to unlock. Ari Patrinos and John Sulston present memoirs of their roles in the HGP, and other key players reminisce on the joys and tribulations that the project bought them –ranging from Yoshiyuki Sakaki's meeting with the Japanese Prime Minister to Fred Sanger again having the last word. The chromosomes are presented as autosomes 1–22 followed by the sex chromosomes, X and Y, rather than chronologically by order of publication. Some of the individual papers are published alongside News and Views articles that introduce the basic work, and provide context and accessibility for key findings. These have been contributed by prominent independent researchers, with the expertise to comment on the relevance and impact of the research.

It has been Nature's privilege to be given the opportunity to work with so many dedicated scientists over the years, particularly in the case of Chris Gunter, who as Nature's current Genetics Editor has collaborated closely with many of these researchers. We are pleased to acknowledge the support of Applied Biosystems, the National Human Genome Research Institute, the Department of Energy and the Wellcome Trust in producing this supplement. The entire contents of the supplement are also available free online (http://www.nature.com/nature/focus/humangenome).

As always, 'Nature' carries sole responsibility for all editorial content and peer review.

10.2 Análisis de datos
10.2.1 Análisis de patrones sintácticos evaluativos

10.2.1.1 Artículos de investigación científica
10.2.1.1.1 Taxonomía de Hunston y Sinclair (2001)

1.1 GRUPO NOMINAL ('IT') + VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA NOMINAL FINITA / NO FINITA
[GN ('IT') + VR + GAdj + Cl] [O Compl]

Categoría evaluativa			Referente evaluado
'It'	Verbo Relacional	Grupo Adjetivo	Cláusula Nominal finita / no finita
It	*seems*	*appropriate* […]	*to briefly summarize two major points of interest concerning isochores.* (T. 10.1.1, lín. 67)
[…] *it*	*is*	*difficult*	*to assess the exact correspondence of the telemetric region described in Caron et al. (2001) with those described in Saccone et al. (1999).* (T. 10.1.3., lín. 83)
It	*is*	*difficult*	*to compare the magnitude of the CpG effect observed here directly to studies of nonplant taxa since methodologies differ* […] (T. 10.1.5., lín. 76)

10.2.1.1.2 Taxonomía de Prieto y Zenteno

1.2 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GN + GV + GN] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal
[…] *the most remarkable ones*	*being*	*the correlations of isochore families* […] (T. 10.1.1, lín. 77)
This [the greatest disparity between TAGs and the genomes compared with the distances among genes in convergent and divergent orientations]	*raises*	*the question of the evolutionary significance of parallel orientations in TAGs* […] (T. 10.1.2, lín. 124)
[…] *genes located in the GC-rich isochores*	*required*	*highly efficient translation* […] (T. 10.1.3, lín. 18)
Re-examination of experimental and theoretical data from publications spanning almost 10 years […]	*allowed us to refine*	*the gene partition criteria first used by Mouchiroud et al., (1991).* (T. 10.1.3, lín. 45)
The last point	*deserves*	*some comments.* (T. 10.1.3, lín. 59)

[…] *our present results*	*support*	*the view that the GC level of the genes is correlated with their expression patterns* […] (T. 10.1.3, lín. 67)
[…] *we*	*did not see*	*a clear relationship between mutation and the composition of individual neighboring nucleotides that do not flank the mutation.* (T. 10.1.5, lín. 102)
[…] *the human SNP study*	*confounded*	*immediate flanking base effects and nonrandom dinucleotide composition.* (T. 10.1.5, lín. 108)
The possibility of a transcription-coupled repair mechanism	*has*	*significant implications for our understanding of compositional bias in genes* […] (T. 10.1.5, lín. 121)

1.3 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN + GV + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal	Frase Preposicional
These findings [ultracentrifugation in Cs2SO4 density gradients in the presence of sequence-specific ligands]	*opened*	*a new inroad*	*in the study of the organization of eukaryotic genomes, superseding DNA reassociation kinetics* […] (T. 10.1.1, lín. 4)

1.4 GRUPO NOMINAL + GRUPO VERBAL + FRASE PREPOSICIONAL
[GN + GV + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Frase Preposicional
Localizing genes in separate isochores	*led*	*to the discovery of an unexpected and strikingly nonrandom distribution of genes* […] (T. 10.1.1, lín. 70)
The research on TAGs	*has been largely confined*	*to individual families of TAGs that serve important physiological functions* […] (T. 10.1.2, lín. 56)

1.5 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL + FRASE PREPOSICIONAL
[FP + GN + GV + FP] [O Simp]

Frase Preposicional	Grupo Nominal	Grupo Verbal	Frase Preposicional
In the last decade,	[…] *several results based on theoretical and experimental approaches*	*led*	*to an improvement of the gene partition criteria* […] (T. 10.1.3, lín. 37)

1.6 GRUPO NOMINAL + GRUPO VERBAL + GRUPO ADVERBIAL
+ CLAÚSULA ADVERBIAL NO FINITA
[GN + GV + GAdv + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal	Grupo Adverbial	Cláusula Adverbial no finita
[…] *they [the findings reported here]*	*go*	*much farther*	*in that they directly identify and map isochores on chromosomes, thus leading to a resolution of >3000 chromosomal bands.* (T. 10.1.1, lín. 31)

1.7 GRUPO NOMINAL + CLÁUSULA ADVERBIAL NO FINITA + GRUPO VERBAL
+ GRUPO NOMINAL + GRUPO ADVERBIAL + CLÁUSULA PREPOSICIONAL
[GN + Cl Adv no fin + GV + GN + GAdv + Cl prep] [O Simp]

Grupo Nominal	Cláusula Adverbial no finita	Grupo Verbal	Grupo Nominal	Grupo Adverbial	Cláusula Preposicional
The present findings,	*while confirming the isochore features previously established,*	*push*	*our knowledge*	*farther*	*by quantifying the size, GC levels, standard deviations, and coordinates of isochores on the human genome map.* (T. 10.1.1, lín. 61)

1.8 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO NOMINAL
[GN + GVpas + GN] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Grupo Nominal
Isochores […]	*have been considered*	*"a fundamental level of genome organization"* […] (T. 10.1.1, lín. 55)

1.9 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA ADVERBIAL
NO FINITA
[GN + GVpas + CL Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal pasivo	Cláusula Adverbial no finita
More studies	*are needed*	*to investigate the underlying mechanism and the nature of this phenomenon.* (T. 10.1.2, lín. 130)

1.10 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL
[GN + GVpas + GAdv] [Cl no fin]

Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial
Gene expression	*is [controlled]*	*more rigidly in germ-line cells than in the surrounding somatic cells.* (T. 10.1.4, lín. 65)
Construcción original: *Gene expression is more rigidly controlled in germ-line cells than in the surrounding somatic cells.*		

1.11 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GN + GVpas + FP] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Frase Preposicional
[…] *the genomes of warm-blooded vertebrates*	*were characterized*	*by a striking long-range compositional heterogeneity* […] (T. 10.1.1, lín. 8)

1.12 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVpas + CL Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal pasivo	Cláusula Nominal no finita
[…] *localized duplication of genomic segments and rearrangement of chromosomal segments*	*have been proposed*	*to be 2 major factors in eukaryotic genome evolution* […] (T. 10.1.2, lín. 8)

1.13 GRUPO NOMINAL + GRUPO VERBAL PASIVO + CLÁUSULA NOMINAL NO FINITA+ CLÁUSULA ADVERBIAL FINITA
[GN + GVpas + Cl Nom no fin + Cl Adv fin] [O Compl]

Grupo Nominal	Grupo Verbal Pasivo	Cláusula Nominal no finita	Cláusula Adverbial finita
The different computational approaches used to disprove or redefine isochores […]	*were* […] *shown*	*to be inadequate* […]	*even if some of them led to a partial identification of isochores.* (T. 10.1.1, lín. 19)

1.14 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GN + GVrel + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
[…] *"such a selective pressure [the fact that human genome organization and evolution are under selective forces] without apparent correlation with gene expression*	*appeared*	*quite speculative".* (T. 10.1.3, lin. 64)
The apparent rarity of retrotransportation in 'cis' (as opposed to between-cell reinfection)	*is* […]	*surprising* […] (T. 10.1.4, lín. 63)
The similar relationship between context and mutation properties in both nuclear and cpDNA	*is*	*interesting* […] (T. 10.1.5, lín. 93)

1.15 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + FRASE PREPOSICIONAL

[GN + GVrel + GAdj + FP] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Frase Preposicional
DNA duplication […]	*is*	*important*	*in adaptive evolution* […] (T. 10.1.2, lín. 1)
Locating on different strands	*is*	*detrimental*	*to the stability of the array* […] (T. 10.1.2, lín. 31)
[…] *shared replication and/or repair processes or that these properties*	*are*	*fundamental*	*to mutations* […] (T. 10.1.5, lín. 94)

1.16 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + CLÁUSULA ADVERBIAL NO FINITA

[GN + GVrel + GAdj + Cl Adv no fin] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Cláusula Adverbial no finita
The apparent rarity of complementation in 'trans'	*may seem*	*surprising*	*given our understanding of the mechanism of retroviral replication.* (T. 10.1.4, lín. 60)

1.17 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO (COMPARATIVO) + CLÁUSULA ADVERBIAL NO FINITA

GN + GVrel + GAdj (Comp) + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo (Comparativo)	Cláusula Adverbial no finita
[…] *parallel orientation in TAGs*	*appears to be*	*more favored than divergent or convergent orientations* […]	*corroborating Graham's conjecture, at least in the 3 genomes.* (T. 10.1.2, lín. 117)

1.18 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL

[GN + GVrel + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
[…] *TAGs*	*are*	*a major component of the genome.* (T. 10.1.2, lín. 41)
[…] *tandem duplication*	*is*	*a major method of gene duplication in many genomes.* (T. 10.1.2, lín. 47)
The common preference for pericentromeric regions by tandem duplication and recent segmental duplication	*might not be*	*a coincidence.* (T. 10.1.2, lín. 93)

An interesting question relevant to the TAG distribution	*is*	*whether the regions that are enriched with TAGs are also rich in other non-TAG duplicates.* (T. 10.1.2, lín. 95)
The hypothesis	*was*	*one of the key points supporting that the human genome organization and evolution are under selective forces* […] (T. 10.1.3, lín. 61)
The original hypothesis of Bernardi (1993)	[…] *was*	*fundamentally correct.* (T. 10.1.3, lín. 94)
[…] *the most apparent neighboring nucleotide effect* […]	*is*	*the CpG effect* […] (T. 10.1.5, lín. 21)
The most notable context effect	*is*	*an elevated rate of CG→TG and CG→CA transitions relative to other transitions.* (T. 10.1.5, lín. 73)

1.19 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN + GVrel + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Frase Preposicional
DNA duplication	[…] *is*	*the principle process patterns* […]	*by which the genetic raw material is provided for the origin of evolutionary novelties such as new gene function and expression* […] (T. 10.1.2, lín. 1)
This [to determine how many duplicated genes are in tandem arrays]	*will*	*shed light*	*on the contribution of tándem duplication to gene duplication in the 3 mammalian genomes.* (T. 10.1.2, lín. 25)

1.20 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + GRUPO ADJETIVO + CLÁUSULA ADVERBIAL / (CLÁUSULA PREPOSICIONAL)
[GN + GVact + GN + GAdj + Cl Adv / (Cl prep)] [O Compl]

Grupo Nominal	Grupo Verbal Activo	Grupo Nominal	Grupo Adjetivo	Cláusula Adverbial / (Cláusula Preposicional)
This recent activity [the HERV-K family first integrated into the genome of the common ancestor of humans and Old World monkeys at least 30 million years ago]	*makes*	*this family*	*ideal*	*for distinguishing between the alternative mechanisms of proliferation.* (T. 10.1.4, lín. 40)

1.21 GRUPO NOMINAL+ GRUPO VERBAL ACTIVO + GRUPO ADJETIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + GAdj + Cl Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Adjetivo	Cláusula Nominal no finita
The availability of SNP data from many different taxa	[…] *makes* […[it]…]	*feasible*	*to develop a more detailed knowledge of factors that contribute to variation in mutational biases.* (T. 10.1.5, lín. 10)
[…] *differences in methodology*	*make* [...[it]...]	*difficult*	*to draw any specific conclusions about differences in context effects.* (T. 10.1.5, lín. 105)

1.22 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + Cl Nom no fin] [O Comp]

Grupo Nominal	Grupo Verbal activo	Cláusula Nominal no finita
[…] *unequal recombination [of a DNA segment]*	[…] *would cause*	*arrays with oppositely oriented repeats to undergo disastrous duplication-deletion events* […] (T. 10.1.2, lín. 114)

1.23 GRUPO NOMINAL + GRUPO ADVERBIAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[GN + GAdv + GVact + GN] [O Simp]

Grupo Nominal	Grupo Adverbial	Grupo Verbal activo	Grupo Nominal
The paucity of inherited stop codons, and the low dN/dS ratios for all genes (including 'env') within the internal branches of the HERV-K (HML2) phylogeny,	*strongly*	*indicate*	*purifying selection* […] (T. 10.1.4, lín. 49)
Continuous selection	*strongly*	*suggests*	*continuous functionality* [...] (T. 10.1.4, lín. 72)

1.24 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GAdv + GN + GVrel + GAdj] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
Surprisingly,	*The mechanism by which HERVs have increased in copy number*	*is*	*only poor understood* (T. 10.1.4, lín. 17)

1.25 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
[GAdv + GN + GVrel + GN] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
Unexpectedly, […]	*regions of increased gene dense*	*are not*	*gene expression* (T. 10.1.3, lín. 75)

1.26 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GAdv + GN + GV + GN] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal	Grupo Nominal
Interestingly,	*the chromosomes that have greater than expected number of TAG forests*	*tend to have*	*less than expected number of TAGs deserts.* (T. 10.1.2, lín. 84)
Unfortunately,	*the figures representing the RIDGEs distribution (Caron et al., 2001)*	*did not allow*	*a precise localization of those regions along the chromosomes.* (T. 10.1.3, lín. 79)

1.27 GRUPO ADVERBIAL + GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL + FRASE PREPOSICIONAL
[GAdv + GN + GVpas + GAdv + FP] [O Simp]

Grupo Adverbial	Grupo Nominal	Grupo Verbal Pasivo	Grupo Adverbial	Frase Preposicional
Until now,	*our understanding of the evolution of interspersed repeats such as HERVs*	*has been influenced*	*heavily*	*by phylogenetic tree shape* […] (T. 10.1.4, lín. 115)

10.2.1.2 Artículos periodísticos

10.2.1.2.1 Taxonomía de Hunston y Sinclair (2001)

1.1 GRUPO NOMINAL ('IT') + VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA
[GN ('IT') + VR + GAdj + Cl] [O Compl]

Categoría Evaluativa Referente Evaluado			
It	Verbo Relacional	Grupo Adjetivo	Cláusula Nominal finita / no finita
[…] *it*	*is*	*too early*	*to tell if AML has different kinds of mutations in different patients.* (T. 10.2.1, lín. 38)
"It	*'s*	*fun*	*to speculate"* […] (T. 10.2.1, lín. 43)
It	*'s*	*a little easier*	*to contrast the macaque with either chimp orhuman to make sense of what's going on.* (T. 10.2.2, lín. 25)

10.2.1.2.2 Taxonomía de Prieto y Zenteno

1.2 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GN + GV + GN] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal
Kevin Shannon […]	*calls*	*this work "a major achievement"* (T. 10.2.1, lín. 23)
They	*have*	*some significant differences among their genes.* (T. 10.2.2., lín. 7)
They [scientists]	*say will enhance*	*medical research in a wide range of areas, including HIV and neuroscience.* (T. 10.2.2., lín. 2)
The findings	*will* [...] *advance*	*scientists' understanding of primate evolution and what makes humans genetically distinct.* (T. 10.2.2., lín. 3)
Tiny slice of genome	*reveals*	*bustling activity in the gaps between genes.* (T. 10.2.3., lín. 1)
The first results from a massive project to exhaustively catalogue all the functions of the human genome	*reveal*	*a hotbed of activity in the gaps between genes.* (T. 10.2.3., lín. 2)
[…] *the significance of which*	*remains*	*a mistery.* (T. 10.2.4., lín. 20)
Proteins	*are*	*the beams and rafters of the cell and the glue that binds the body together; they're the hormones that course through our veins and the guided missiles that target infections; they're the enzymes that build up and break down our energy reserves and the circuits that power movement and thought.* (T. 10.2.5., lín. 8)
The flurry of private activity	*raises*	*the specter of intellectual-property disputes like those that plagued the Human Genome Project.* (T. 10.2.5., lín. 46)

1.3 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
+ FRASE PREPOSICIONAL
[GN + GV + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal	Frase Preposicional
[…] *scientists*	*began*	*the next and far more challenging step*	*in explaining the molecular underpinnings of life.* (T. 10.2.5., lín. 1)

1.11 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GN + GVpas + FP] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Frase Preposicional
[…] *this genetic culprit*	*was revealed*	*by a second and little-heralded phase in the Human Genome Project* […] (T. 10.2.3., lín. 9)

1.14 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GN + GVrel + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
[…] *this newfound ability to determine the complete DNA sequence of a cancer cell*	*"is*	*enormously powerful"* […] (T. 10.2.1., lín. 7)
We	*were*	*flying blind* […] (T. 10.2.1., lín. 29)
[…] *individual AML cases*	*are*	*distinct.* (T. 10.2.1., lín. 33)
It [population genetics]	*is*	*seldom predictive.* (T. 10.2.3., lín. 80)

1.15 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
+ FRASE PREPOSICIONAL
[GN + GVrel + GAdj + FP] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Frase Preposicional
[…] *it*	*'s*	*not true*	*for most common diseases.* (T. 10.2.3., lín. 89)
[…] *a lot of people*	*were*	*skeptical*	*of how real intergenic activity was* […] (T. 10.2.4., lín. 38)

1.16 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
+ CLÁUSULA ADVERBIAL NO FINITA
[GN + GVrel + GAdj + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal Relacional	Grupo Adjetivo	Cláusula Infinitiva	Cláusula Adverbial no finita
These alterations [Tay- Sachs disease and Huntington disease]	*are*	*relatively easy*	*to identify*	*because they can be traced and isolated in families with a history of the disease.* (T. 10.2.3., lín. 14)

1.17 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO (COMPARATIVO) + CLAÚSULA ADVERBIAL NO FINITA
[GN + GVrel + GAdj (comp) + Cl Adv no fin] [O Compl]

Grupo Nominal	Grupo Verbal Relacional	Grupo Adjetivo (comparativo)	Cláusula Adverbial no finita
Finding genetic clues to common diseases	*is*	*much more difficult*	*because many genes-as well as lifestyle and environmental exposures- may be involved.* (T. 10.2.3., lín. 15)

1.18 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
[GN + GVrel + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
[…] *here*	*are*	*your best treatment options*. (T. 10.2.1., lín. 48)
This [the fact that the chimpanzee and the macaque share a feature and the human is different]	*is*	*a human change*. (T. 10.2.2., lín. 27)
Some interesting findings	*are*	*already coming from a study of variations within the macaque genome.* (T. 10.2.2., lín. 51)
The project, called the HapMap,	*is*	*the effort of a six nation consortium.* (T. 10.2.3., lín. 26)
There	*are*	*a lot of hucksters out there saying that this or that gene will predict a disease- it's not true for most common diseases.* (T. 10.2.3., lín. 88)
[…] *it [comparing people with a disease to those without the disease]*	*is not*	*useful information*. (T. 10.2.3., lín. 87)
This [the ENCODE project]	*is*	*a 'community resource project', like genome projects, that makes lots of new data available to the community* […] (T. 10.2.4., lín. 76)
That	*'s*	*a tiny fraction of all the proteins found in nature.* (T. 10.2.5., lín. 18)

1.20 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + GRUPO ADJETIVO + CLÁUSULA ADVERBIAL / (CLÁUSULA PREPOSICIONAL)
[GN + GVact + GN + GAdj + Cl Adv / (CL prep)] [O Simp]

Grupo Nominal	Grupo Verbal Activo	Grupo Nominal	Grupo Adjetivo	Cláusula Adverbial / (Cláusula Preposicional)
We	*'re knowing*	*the macaque*	*better*	*instead of just advocating experimenting on the macaque.* (T. 10.2.2., lín. 81)

1.21 GRUPO NOMINAL+ GRUPO VERBAL ACTIVO + GRUPO ADJETIVO + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + GAdj + Cl Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Adjetivo	Cláusula Nominal no finita
This new sequencing technology, called massively parallel sequencing,	*makes* […[it]…]	*posible*	*to compare the normal DNA sequence to the cancerous DNA sequence in the same patient.* (T. 10.2.1., lín. 18)

1.28 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO (COMPARATIVO)
[GN + GVrel + GAdj (Comp)] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo (Comparativo)
No drugs	*are*	*effective for the other, far more common form of the disease.* (T. 10.2.3., lín. 48)
I	*'m*	*less interested in tests than treatments* (T. 10.2.3., lín. 103)
[…] *they [proteins]*	*are*	*far more variegated and complex than DNA.* (T. 10.2.5., lín. 6)

1.29 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + CLÁUSULA NOMINAL
[GN + GVrel + Cl Nom] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Cláusula Nominal
[…] *our hopes*	*are*	*to do more.* (T. 10.2.1, lín. 61)

1.30 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + GRUPO VERBAL ACTIVO + FRASE PREPOSICIONAL
[GN + GVrel + GN + GVact + FP) [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Grupo Verbal activo	Frase Preposicional
[…] *there*	*was*	*interesting stuff*	*going on*	*in these regions [between genes]* […] (T. 10.2.4., lín. 9)

1.31 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + FRASE PREPOSICIONAL
[GN + GVact + FP] [O Simp]

Grupo Nominal	Grupo Verbal activo	Frase Preposicional
No cure or effective treatment	*exists*	*for Alzheimer's* […] (T. 10.2.3., lín. 52)

1.32 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + CLÁUSULA NOMINAL NO FINITA
[GN + GVact + GN + Cl Nom no fin] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Cláusula Nominal no finita
Scientists	*expect*	*the rhesus macaque genome sequence*	*to enhance research in neuroscience, behavioral biology, reproductive physiology, endocrinology, and cardiovascular studies.* (T. 10.2.2., lín. 47)

1.33 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL PRONOMINAL + GRUPO ADJETIVO
[GN + GVact + GN (pro) + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal (pronominal)	Grupo Adjetivo
[…] *we*	*'re doing*	*something*	*quite opposite* […] (T. 10.2.2., lín. 79)

1.34 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + GRUPO ADJETIVO + FRASE PREPOSICIONAL + CLÁUSULA ADVERBIAL FINITA. / (CLÁUSULA PREPOSICIONAL)
[GN + GVact + GN + GAdj + FP + Cl Adv fin / (Cl prep) [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Grupo Adjetivo	Frase Preposicional	Cláusula Adverbial finita / (Cláusula preposicional)
That divergence	*makes*	*macaques*	*ideal*	*for the evolutionary study of primates,*	*because important features that have been conserved in primates over time can be more easily seen by comparing rhesus monkeys to humans than by comparing chimps to humans.* (T. 10.2.2., lín. 19)

1.35 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + CLÁUSULA INFINITIVA + CLÁUSULA ADVERBIAL /(CLAUSULA PREPOSICIONAL)
[GN + GVact + GN + Cl inf + Cl Adv / (Cl prep)] [O Compl]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Cláusula Infinitiva	Cláusula Adverbial / Cláusula Preposicional
[…] *it [5 percent of the studied DNA sequence]*	*plays*	*an <u>important enough role</u>*	*for evolution to preserve*	*while species have evolved.* (T. 10.2.4., lín. 64)

1.36 GRUPO ADVERBIAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GAdv + GVpas + FP] [O Simp]

Grupo Adverbial	Grupo Verbal pasivo	Frase Preposicional
<u>Little</u>	*<u>is known</u>*	*about the differences between a male and a female genome.* (T. 10.2.1., lín. 56)

1.37 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PRONOMINAL (PREPOSICIONAL) + FRASE PREPOSICIONAL
[GN + GVpas + FPro (prep) + FP] [O Simp]

Grupo Nominal	Grupo Verbal Pasivo	Frase Pronominal (preposicional)	Frase Preposicional
[…] *the gene*	*is situated*	*<u>somewhere</u>*	*along a particular DNA strand between two easily recognized sequences called markers* […] (T. 10.2.2., lín. 70)

1.38 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + GRUPO ADJETIVO (COMPARATIVO)
[GN + GVrel + GN + GAdj (comp)] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Grupo Adjetivo (comparativo)
[…] *the two subspecies of the macaques*	*are*	*very different from each other on a genetic level,*	*<u>probably much more different</u> <u>than</u> human populations are from each other* (T. 10.2.2., lín. 58)

1.39 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL + GRUPO NOMINAL
[GN + GVpas + GAdv + GN] [O Simp]

Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial	Grupo Nominal
[…] *robots and powerful X-ray generators*	*<u>have boosted</u>*	*[<u>lately</u>]*	*the pace of discovery.* (T. 10.2.5., lín. 37)
Construcción original: […] *robots and powerful X-ray generators <u>have lately boosted</u> the pace of discovery.*			

1.40 GRUPO NOMINAL (PRONOMINAL) + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN (PRO) + GVrel + GN + FP] [O Simp]

Grupo Nominal (pronominal)	Grupo Verbal relacional	Grupo Nominal	Frase Preposicional
[…] *there*	*is*	*good news*	*for some patients.* (T. 10.2.3., lín. 44)

1.41 CLÁUSULA ADVERBIAL NO FINITA + GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[Cl Adv no fin + GN + GVact + GN] [O Compl]

Cláusula Adverbial no finita	Grupo Nominal	Grupo Verbal activo	Grupo Nominal
[…] *after rigorously pruning the data to keep only the most significant mutations,*	*the researchers*	*identified*	*10 mutations* […] (T. 10.2.1., lín. 29)

10.2.1.3 Editoriales

10.2.1.3.1 Taxonomía de Hunston y Sinclair (2001)

1.1 GRUPO NOMINAL ('IT') + VERBO RELACIONAL + GRUPO ADJETIVO + CLÁUSULA [GN ('IT') + VR + GAdj + Cl] [O Compl]

		Categoría Evaluativa	Referente Evaluado
'It'	Verbo Relacional	Grupo Adjetivo / Grupo Nominal	Cláusula Nominal finita / no finita
[…] *it*	*is*	*hard*	*to see how that investment could have received better returns.* (T. 10.3.2., lín. 10)
[…] *it*	*is*	*encouraging*	*that in the UK there were no problems about confidentiality* […] (T. 10.3.3., lín. 44)
[…] *families*	*were*	*confident*	*that the register would be used in a positive and not in a discriminatory way.* (T. 10.3.3., lín. 44)
It	*is not*	*inconceivable*	*that this will happen again, only it will be for medicine* (T. 10.3.4., lín. 37)
It	*was*	*clear*	*that they could not be coordinated in the same way.* (T. 10.3.5., lín. 29)
It	*was*	*important*	*that each team had the time to complete their analysis to the highest level and to produce the best possible product without the work being lost in a competitive race to publish first* […] (T. 10.3.5., lín. 33)

10.2.1.3.2 Taxonomía de Prieto y Zenteno

1.2 GRUPO NOMINAL + GRUPO VERBAL + GRUPO NOMINAL
[GN + GV + GN] [O Simp]

Grupo Nominal	Grupo Verbal	Grupo Nominal
[…] *The transcription part of the problem*	*shifted*	[…] *to <u>the knotty problem</u> of how damage is repaired* […] (T. 10.3.1., lín. 23)
Translation	*became*	*<u>the key issue</u>* […] (T. 10.3.1., lín.28)
To do the same in eukaryotic cells	*proved*	*a <u>more difficult challenge</u>.* (T. 10.3.1., lín.40)

1.4 GRUPO NOMINAL + GRUPO VERBAL + FRASE PREPOSICIONAL
[GN + GV + FP] [O Simp]

Grupo Nominal	Grupo Verbal	Frase Preposicional
This issue of 'Science'	*appears*	*<u>amid a swirl of contemporary reminiscence</u> about DNA* […] (T. 10.3.1., lín. 1)

1.8 GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO NOMINAL
[GN + GVpas + GN] [O Simp]

Grupo Nominal	Grupo Verbal Pasivo	Grupo Nominal
In keeping with the spirit of the genomics community work ethics, 'Nature'.	*has made*	*each of these papers <u>freely available</u> on the Internet* (T. 10.3.5., lín. 20)

1.11 GRUPO NOMINAL + GRUPO VERBAL PASIVO + FRASE PREPOSICIONAL
[GN + GVpas + FP] [O Simp]

Grupo Nominal	Grupo Verbal Pasivo	Frase Preposicional
This <u>seminal work</u>	*has* […] *been followed*	*by the publication of the remaining chromosomes* […] (T. 10.3.5., lín. 18)

1.14 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
[GN + GVrel + GAdj] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo
Making freely available all the materials used for a piece of research	*is*	*<u>sometimes impractical</u>.* (T.10.3.2., lín. 26)
The amount of valuable scientific information residing in research and development centres in the private sector	*is*	*<u>huge</u>* […] (T.10.3.2., lín. 59)

1.18 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL
[GN + GVrel + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal
[…] *the human genome in particular*	*is* […]	*humanity's common heritage*. (T.10.3.2., lín. 23)
There	*are*	*too many challenges around for it to make sense for public and private scientists to duplicate one another's work* (T.10.3.2., lín. 54)
('Personal contact	*is*	*the key to success')* (T.10.3.3., lín. 41)
The newspapers	*were*	*full of headlines screaming of the benefits of the project to humanity.* (T.10.3.4., lín. 3)
A most distressing feature that has come to the fore	*is*	*the facility with which we tout our large (almost impossibly so) population and ethnic diversity as a major resource in biological research* […] (T.10.3.4., lín. 63)

1.19 GRUPO NOMINAL (PRONOMINAL) + GRUPO VERBAL RELACIONAL
+ GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN (Pro) + GVrel + GN + FP] [O Simp]

Grupo Nominal (Pronominal)	Grupo Verbal relacional	Grupo Nominal	Frase Preposicional
There	*is*	*big money*	*at stake.* (T. 10.3.2., lín. 32)

1.23 GRUPO NOMINAL + GRUPO ADVERBIAL + GRUPO VERBAL ACTIVO
+ GRUPO NOMINAL
[GN + GAdv + GVact + GN] [O Simp]

Grupo Nominal	Grupo Adverbial	Grupo Verbal activo	Grupo Nominal
It [the human genome sequence]	*profoundly*	*influences*	*our bodies, our behaviour and our minds* […] (T. 10.3.2., lín. 5)
'Nature'	*fully*	*appreciated*	*the scientific value of publishing all 23 pairs of chromosomes as a single entity.* (T. 10.3.5., lín. 36)

1.31 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + FRASE PREPOSICIONAL
[GN + GVact + FP] [O Simp]

Grupo Nominal	Grupo Verbal activo	Frase Preposicional
Good public health […]	*rests*	*on empirical data of high quality.* (T. 10.3.3., lín. 1)
[…] *our repeated drum beating on information technology (IT)*	*rests*	*on the fragile fabric of a non-existent hardware base.* (T. 10.3.4., lín. 77)

1.42 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL + FRASE PREPOSICIONAL
[GN + GVact + GN + FP] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal	Frase Preposicional
The burgeoning commercial sector that is based on genome information	*poses*	*a challenge*	*to the norms of scientific publication.* (T. 10.3.2., lín. 1)

1.43 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL PASIVO + GRUPO ADVERBIAL
[FP + GN + GVpas + GAdv] [O Simp]

Frase Preposicional	Grupo Nominal	Grupo Verbal pasivo	Grupo Adverbial
With this shift,	*the burden of genetic disease*	*has changed*	*dramatically.* (T. 10.3.3., lín. 14)

1.44 FRASE PREPOSICIONAL + GRUPO NOMINAL + GRUPO VERBAL ACTIVO + CLÁUSULA INFINITIVA
[FP + GN + GVact + Cl Adv no fin] [O Compl]

Frase Preposicional	Grupo Nominal	Grupo Verbal activo	Cláusula Adverbial no finita
From a purely scientific standpoint,	*biology research*	*stands*	*to benefit enormously from the availability of the genome sequences of diverse organisms.* (T. 10.3.4., lín. 43)

1.45 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + GRUPO NOMINAL
[GN + GVrel + GAdj + GN] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Grupo Nominal
'Nature'	*is*	*delighted*	*this week to publish the project's analysis.* (T. 10.3.2., lín. 12)
The private sector	*is*	*essential*	*to the improvement of human health* […] (T. 10.3.2., lín. 56)

1.46 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO + CLÁUSULA INFINITIVA
[GN + GVrel + GAdj + Cl inf] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Cláusula Infinitiva
[…] *it*	*is*	*unethical*	*for a country to push families towards screening* (T. 10.3.3., lín. 54)
[…] *it*	*is*	*equally unethical*	*to deprive a couple of that information* […] (T. 10.3.3., lín. 55)

1.47 GRUPO NOMINAL (PRONOMINAL) + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + GRUPO ADJETIVO + CLÁUSULA INFINITIVA
[GN (Pro) + GVrel + GN + GAdj + Cl inf] [O Compl]

Grupo Nominal	Grupo Verbal Relacional	Grupo Nominal	Grupo Adjetivo	Cláusula Infinitiva
[…] *there*	*was*	*a splendid storehouse of information*	*already available*	*to provide that resonance.* (T. 10.3.1., lín. 6)

1.48 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + CLÁUSULA NOMINAL NO FINITA
[GN + GVrel + GN + Cl nom fin] [O Compl]

Grupo Nominal	Grupo Verbal relacional	Grupo Nominal	Cláusula nominal finita
It	*is*	*a pity*	*that Professor Modell did not comment on this point* […] (T. 10.3.3., lín. 42)
[…] *there*	*is*	*a worrying tendency*	*to be self-congratulatory about our achievements in biological research* […] (T. 10.3.4., lín. 68)

1.49 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO NOMINAL + FRADE PREPOSICIONAL + CLÁUSULA RELATIVA / EXPLICATIVA
[GN + GVrel + GN + FP + Cl rel / expl] [O Simp]

Grupo Nominal	Grupo Verbal Relacional	Grupo Nominal	Frase Preposicional	Cláusula relativa / explicativa
This [the production of the first 'working draft' of the human genome]	*was*	*a landmark*	*in scientific research and a momentous occasion for 'Nature',*	*which was the scientific journal with responsibility for the review and evaluation of the public effort* […] (T. 10.3.5., lín. 4)

1.50 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL
+ CLÁUSULA ADVERBIAL NO FINITA
[GN + GVrel + Cl Adv no fin] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Cláusula Adverbial no finita
The offer	*was*	*for 'Nature' to review the work with the usual rigour bestowed on all of its scientific submissions* […] (T. 10.3.5., lín.39)

1.51 GRUPO NOMINAL + GRUPO VERBAL RELACIONAL + GRUPO ADJETIVO
+ CLÁUSULA INFINITIVA
[GN + GVrel + GAdj + CL inf] [O Simp]

Grupo Nominal	Grupo Verbal relacional	Grupo Adjetivo	Clausula infinitiva
[…] *a claim those authors*	*would be*	*quick*	*to disavow* […] (T. 10.3.1., lín. 8)

1.52 GRUPO NOMINAL + GRUPO VERBAL ACTIVO + GRUPO NOMINAL
[GN + GVact + GN] [O Simp]

Grupo Nominal	Grupo Verbal activo	Grupo Nominal
The collated information from these groups	*would provide*	*the most comprehensive analysis of the human genome.* (T. 10.3.5., lín. 15)

1.53 CLÁUSULA NOMINAL NO FINITA + GRUPO VERBAL RELACIONAL
+ GRUPO NOMINAL
[Cl nom no fin + GVrel + GN] [O Simp]

Cláusula Nominal no finita	Grupo Verbal relacional	Grupo Nominal
To do the same in eukaryotic cells	*proved*	*a more difficult challenge.* (T. 10.3.1., lín. 41)
Evolution	*is*	*ultimately dependent on mutation.* (T. 10.3.5., lín. 1)

10.2.2 Análisis de significados valorativos
10.2.2.1 Artículos de investigación científica

Texto 10.1.1

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(1) lín. 4	*[Ultracentrifugation in Cs2SO4 density gradients in the presence of sequence-specific ligands]*	EVext	*These findings opened a new inroad in the study of the organization of eukaryotic genomes, superseding DNA reassociation kinetics* […]	VT	(+)
(2) lín. 8	[…] *the genomes of warm-blooded vertebrates*	EVext	*were characterized by a striking long-range compositional heterogeneity* […]	VT	(+)
(3) lín. 19	*The different computational approaches used to disprove or redefine isochores* […]	EVext	*were* […] *shown to be inadequate* […] *even if some of them led to a partial identification of isochores.*	VT	(-)
(4) lín. 31	*[The findings reported here]*	EVint	[…] *they go much farther in that they directly identify and map isochores on chromosomes, thus leading to a resolution of >3000 chromosomal bands.*	VT	(+)
(5) lín. 55	*Isochores* […]	EVext	*have been considered "a fundamental level of genome organization"* […]	VT	(+)
(6) lín. 61	*The present findings, while confirming the isochore features previously established,*	EVint	*push our knowledge farther, by quantifying the size, GC levels, standard deviations, and coordinates of isochores on the human genome map.*	VT	(+)
(7) lín. 67	*[Isochores]*	EVint	*It seems appropriate here to briefly summarize two major points of interest concerning isochores.*	VT	(+)
(8) lín. 70	*Localizing genes in separate isochores*	EVext	*led to the discovery of an unexpected and strikingly nonrandom distribution of genes* […]	VT	(N)
(9) lín. 77	*[Several different properties]*	EVext	[…] *the most remarkable ones being the correlations of isochore families* […]	VT	(+)

Texto 10.1.2

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(10) lín. 1	*DNA duplication*	EVext	[…] *is <u>important</u> in adaptive evolution* […]	VT	(N)
(11) lín. 1	*DNA duplication*	EVint	[…] *is <u>the principle process</u> by which the genetic raw material is provided for the origin of evolutionary novelties such as new gene function and expression patterns* […]	VT	(N)
(12) lín. 8	[…] *localized duplication of genomic segments and rearrangement of chromosomal segments*	EVext	*have been proposed to be 2 <u>major factors</u> in eukaryotic genome evolution* […]	VT	(N)
(13) lín. 25	*[To determine how many duplicated genes are in tandem arrays]*	EVint	*This <u>will shed light</u> on the contribution of tandem duplication to gene duplication in the 3 mammalian genomes.*	VT	(+)
(14) lín. 3l	[…] *locating on different strands*	EVint	*is <u>detrimental</u> to the stability of the array* […]	VT	(-)
(15) lín. 41	[…] *TAGs*	EVext	*are a <u>major component</u> of the genome.*	VT	(N)
(16) lín. 47	[…] *tandem duplication*	EVint	*is a <u>major method</u> of gene duplication in many genomes.*	VT	(N)
(17) lín. 56	*The research on TAGs*	EVint	*has been <u>largely</u> confined to individual families of TAGs that serve important physiological functions* […]	VT	(N)
(18) lin. 84	*[Chromosomes that have greater than expected number of TAG forests]*	EVint	*<u>Interestingly</u>, the chromosomes that have greater than expected number of TAG forests tend to have less than expected number of TAGs deserts.*	AFE	(+)
(19) lin. 93	*The common preference for pericentromeric regions by tandem duplication and recent segmental duplication*	EVint	*<u>might not be a coincidence</u>.*	VT	(N)
(20) lin. 95	*[TAG distribution]*	EVint	*An <u>interesting question</u> relevant to the TAG distribution is whether the regions that are enriched with TAGs are also rich in other non-TAG duplicates.*	VT	(N)

(21) lín. 114	[…] *unequal recombination [of a DNA segment]*	EVint	[…] *would cause arrays with oppositely oriented repeats to undergo disastrous duplication-deletion events* […]	VT	(-)
(22) lín. 117	[…] *parallel orientation in TAGs*	EVext	*appears to be more favored than divergent or convergent orientations* […] *corroborating Graham's conjecture, at least in the 3 genomes.*	VT	(+)
(23) lín. 124	*This [the greatest disparity between TAGs and the genomes compared with the distances among genes in convergent and divergent orientations]*	EVint	*raises the question of the evolutionary significance of parallel orientations in TAGs* […]	VT	(N)
(24) lín. 130	*[The effect of parallel versus other types of orientation in TAGs]*	EVint	*More studies are needed to investigate the underlying mechanism and the nature of this phenomenon.*	VT EVimp	(-)

Texto 10.1.3

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(25) lín. 18	[…] *genes located in the GC-rich isochores*	EVint	*required highly efficient translation* […]	VT	(N)
(26) lín. 37	*In the last decade,* […] *several results based on theoretical and experimental approaches*	EVint	*led to an improvement of the gene partition criteria* […]	VT	(+)
(27) lin. 47	*[The gene partition criteria first used by Mouchiroud et al., (1991).]*	EVint	*Re-examination of experimental and theoretical data from publications spanning almost 10 years* […] *allowed us to refine the gene partition criteria first used by Mouchiroud et al., (1991).*	VT	(+)
(28) lín. 59	*[Tissue-specific and widely expressed genes follow the general gene distribution, reaching the highest frequency in the GC-rich isochores]*	EVint	*The last point deserves some comments.*	VT	(+)
(29) lín. 61	*[The gene density and the GC level of isochores were hypothesized to be correlated with the gene expression levels]*	EVint	*The hypothesis was one of the key points supporting that the human genome organization and evolution are under selective forces* […]	VT	(+)
(30) lín. 64	[…] *"such a selective pressure [the fact that the human genome organization and evolution are under selective forces] without apparent correlation with gene expression*	EVext	*appeared quite speculative."*	VT	(-)
(31) lín. 67	[…] *our present results*	EVint	*support the view that the GC level of the genes is correlated with their expression patterns* […]	VT	(+)
(32) lín. 75	*[Regions of increased gene expression are not gene dense]*	EVint	*Unexpectedly,* […]	VT	(-)
(33) lín. 79	*[The figures representing the RIDGEs distribution]*	EVint	*Unfortunately,* […]	VT	(-)
(34) lín. 83	*[The exact correspondence of the telemetric region described in Caron et al.*	EVext	[…] *it is difficult to assess* […]	VT	(-)

	(2001), with those described in Saccone et al. (1999).]				
(35) lín. 94	*The original hypothesis of Bernardi (1993)*	EVint	[…] *was <u>fundamentally correct</u>.*	VT	(+)

Texto 10.1.4

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(36) lín. 17	*[The mechanism by which HERVs have increased in copy number is only poor understood]*	EVint	*Surprisingly*, […]	VT	(N)
(37) lín. 17	*The mechanism by which HERVs have increased in copy number*	EVint	[…] *is only poorly understood.*	VT	(-)
(38) lín. 40	*[The HERV-K family first integrated into the genome of the common ancestor of humans and Old World monkeys at least 30 million years ago]*	EVint	*This recent activity makes this family ideal for distinguishing between the alternative mechanisms of proliferation.*	VT	(+)
(39) lín. 49	*The paucity of inherited stop codons, and the low dN/dS ratios for all genes (including 'env') within the internal branches of the HERV-K (HML2) phylogeny,*	EVint	*strongly indicate purifying selection* […]	VT	(N)
(40) lín. 60	*The apparent rarity of complementation in 'trans'*	EVint	*may seem surprising given our understanding of the mechanism of retroviral replication.*	VT	(N)
(41) lín. 63	*The apparent rarity of retrotransportation in 'cis' (as opposed to between-cell reinfection)*	EVint	*is* […] *surprising* […]	VT	(N)
(42) lín. 65	[…] *gene expression*	EVint	*is more rigidly controlled in germ-line cells than in the surrounding somatic cells.*	VT	(N)
(43) lín. 72	*Continuous selection*	EVint	*strongly suggests continuous functionality* […]	VT	(N)
(44) lín. 115	*Until now, our understanding of the evolution of interspersed repeats such as HERVs*	EVint	*has been influenced heavily by phylogenetic tree shape* […]	VT	(N)

Texto 10.1.5

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(45) lín. 1	*Evolution*	EVint	*is ultimately dependent on mutation.*	VT	(N)
(46) lín. 10	*The availability of SNP data from many different taxa*	EVext	*now makes it feasible to develop a more detailed knowledge of factors that contribute to variation in mutational biases.*	VT	(N)
(47) lín. 21	*In nuclear genes,*	EVint	*The most apparent neighbouring nucleotide effect* […] *is the CpG effect* […]	VT	(N)
(48) lín. 73	*The most notable context effect*	EVint	*is an elevated rate of CG→TG and CG→CA transitions relative to other transitions.*	VT	(N)
(49) lín. 76	*[The magnitude of the CpG effect observed here directly to studies of nonplant taxa]*	EVint	*It is difficult to compare the magnitude of the CpG effect observed here directly to studies of nonplant taxa since methodologies differ* […]	VT	(-)
(50) lín. 93	*The similar relationship between context and mutation properties in both nuclear and cpDNA*	EVint	*is interesting* […]	VT	(N)
(51) lín. 94	[…] *shared replication and/or repair processes or that these properties*	EVint	*are fundamental to mutations* […]	VT	(+)
(52) lín. 102	[…]*We*	EVint	*did not see a clear relationship between mutation and the composition of individual neighboring nucleotides that do not flank the mutation.*	VT	(-)
(53) lín. 105	[…] *differences in methodology*	EVint	*make it difficult to draw any specific conclusions about differences in context effects.*	VT	(-)
(54) lín. 108	[…] *the human SNP study*	EVint	*confounded immediate flanking base effects and nonrandom dinucleotide composition.*	VT	(-)
(55) lín. 121	*The possibility of a transcription-coupled repair mechanism*	EVint	*has significant implications for our understanding of compositional bias in genes* […]	VT	(+)

10.2.1.2 Artículos periodísticos

Texto 10.2.1

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(56) lín. 7	[…] *this newfound ability*	EVext	*to determine the complete DNA sequence of a cancer cell is enormously powerful* […]	VT	(+)
(57) lín. 18	*This new sequencing technology, called massively parallel sequencing,*	EVint	*makes it possible to compare the normal DNA sequence to the cancerous DNA sequence in the same patient.*	VT	(N)
(58) lín. 23	*Kevin Shannon* […]	EVext	*calls this work "a major achievement"*	VT	(+)
(59) lín. 29	*[The number of mutated genes in the cancerous cells]*	EVint	[…] *after rigorously pruning the data to keep only the most significant mutations, the researchers identified 10 mutations* […]	VT	(+)
(60) lín. 29	*We*	EVext	*were flying blind .*	VT	(-)
(61) lín. 33	[…] *individual AML cases*	EVext	*are distinct.*	VT	(N)
(62) lín. 38	*[AML disease]*	EVint	*It is too early to tell if AML has different kinds of mutations in different patients.*	VT	(-)
(63) lín. 43	*[Common mutations in similar groups of genes]*	EVext	*"It's fun to speculate* […]	AFE	(+)
(64) lín. 48	*[Doctors telling their patients about their diseases]*	EVext	[…] *here are your best treatment options.*	VT	(+)
(65) lín. 56	*[Differences between a male and a female genome]*	EVint	*Little is known about the differences between a male and a female genome.*	VT	(N)
(66) lín. 61	*[More studies of cancer genome sequences]*	EVint	[…] *our hopes are to do more.*	AFE EVimp	(+)

Texto 10.2.2

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(67) lín. 2	*They [scientists]*	EVext	*say will enhance medical research in a wide range of areas, including HIV and neuroscience.*	VT	(+)
(68) lín. 3	*The findings*	EVint	*will also advance scientists' understanding of primate evolution and what makes humans genetically distinct.*	VT	(+)
(69) lín. 7	*They*	EVint	*have some significant differences among their genes.*	VT	(N)
(70) lín. 19	*[The chimp genome is only about 1.5 percent different, while the macaque genome is about 7 percent different]*	EVint	*That divergence makes macaques ideal for the evolutionary study of primates, because important features that have been conserved in primates over time can be more easily seen by comparing rhesus monkeys to humans than by comparing chimps to humans.*	VT	(+)
(71) lín. 25	*[The macaque with either chimp or human]*	EVext	*It's a little easier to contrast the macaque with either chimp or human to make sense of what's going on.*	VT	(+)
(72) lín. 27	*This [the fact that the chimpanzee and the macaque share a feature and the human is different]*	EVext	*is a human change.*	VT	(N)
(73) lín. 47	*[The rhesus macaque genome sequence]*	EVext	*Scientists expect the rhesus macaque genome sequence to enhance research in neuroscience, behavioral biology, reproductive physiology, endocrinology, and cardiovascular studies.*	VT	(+)
(74) lín. 51	*Some interesting findings*	EVint	*are already coming from a study of variations within the macaque genome.*	VT	(N)
(75) lín. 58	[…] *the two subspecies of the macaques*	EVint	*are very different from each other on a genetic level, probably much more different than human populations are from each other.*	VT	(N)
(76) lín. 70	[…] *the gene*	EVext	*is situated somewhere along a particular DNA strand between two easily recognized sequences called markers* […]	VT	(N)

(77) lín. 79	*[The number of macaques used in laboratory research]*	EVint	[…] *we're doing <u>something quite opposite</u>,"* […]	VT	(N)
(78) lín. 81	*[The rhesus macaque]*	EVint	*<u>We're knowing the macaque better</u> instead of just advocating experimenting on the macaque.*	VT	(+)

Texto 10.2.3

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(79) lín. 9	[…] *this genetic culprit*	EVint	*was revealed by a second and little-heralded phase in the Human Genome Project* […] *T. 10.2.3*	AFE	(-)
(80) lín. 14	*[Tay-Sachs disease and Huntington disease]*	EVint	*These alterations are relatively easy to identify because they can be traced and isolated in families with a history of the disease.*	VT	(+)
(81) lín. 15	*Finding genetic clues to common diseases*	EVint	*is much more difficult because many genes-as well as lifestyle and environmental exposures-may be involved.*	VT	(-)
(82) lín. 26	*The project, called the HapMap,*	EVint	*is the effort of a six nation consortium*	AFE	(+)
(83) lín. 44	*[The drug Lucentis stabilizing or even improving vision]*	EVint	[…] *there is good news for some patients.*	AFE	(+)
(84) lín. 48	*No drugs*	EVint	*are effective for the other, far more common form of the disease.*	VT	(-)
(85) lín. 52	*[A cure or treatment s for Alzheimer's]*	EVint	*No cure or effective treatment exists for Alzheimer's* […] *T.10.2.3.*	AFE	(-)
(86) lín. 80	*[The study of a gene or genes responsible for a disease]*	EVint	*It is seldom predictive.*	VT	(N)
(87) lín. 87	*[Comparing people with a disease to those without the disease]*	EVint	[…] *it is not useful information.*	VT	(-)
(88) lín. 88	*[Those geneticists reporting that this or that gene will predict a disease]*	EVint	*There are a lot of hucksters out there saying that this or that gene will predict a disease-it's not true for most common diseases.*	AFE	(-)
(89) lín. 89	[…] *it [this or that gene will predict a disease]*	EVint	*it's not true for most common diseases.*	VT	(-)
(90) lín. 103	*[Tests for genes responsible for diseases and treatments for diseases]*	EVext	*I'm less interested in tests than treatments.*	AFE	(-)

Texto 10.2.4

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría valorativa	Valor (+) / (-) / (N)
(91) lín. 1	*Tiny slice of genome*	EVint	*reveals bustling activity in the gaps between genes.*	VT	(N)
(92) lín. 2	*The first results from a massive project to exhaustively catalogue all the functions of the human genome*	EVint	*reveal a hotbed of activity in the gaps between genes.*	VT	(+)
(93) lín. 9	*[The gaps between genes in the human genome sequence]*	EVext	[…] *there was interesting stuff going on in these regions [between genes]* […]	VT	(N)
(94) lín. 20	[…] *the significance of which [the publication of the human genome sequence in 2001]*	EVint	*remains a mistery.*	VT	(N)
(95) lín. 23	*They [Noncoding regions of the human DNA sequence]*	EVint	*play an important function.*	VT	(N)
(96) lín. 38	[…] *a lot of people*	EVint	*were skeptical of how real intergenic activity was* […]	VT	(-)
(97) lín. 64	*[5 percent of the studied DNA sequence]*	EVint	[…] *it plays an important enough role for evolution to preserve while species have evolved.*	VT	(N)
(98) lín. 76	*This [the ENCODE project]*	EVext	*is a 'community resource project', like genome projects, that makes lots of new data available to the community* […]	VT	(+)

Texto 10.2.5

Número / línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(99) lín. 1	*[To explain the molecular underpinnings of life]*	EVint	[…] *scientists began the next and far more challenging step in explaining the molecular underpinnings of life.*	AFE	(+)
(100) lín. 6	*[Proteins in the human body]*	EVint	[…] *they are far more variegated and complex than DNA.*	VT	(N)
(101) lín. 8	*Proteins*	EVint	*are the beams and rafters of the cell and the glue that binds the body together; they're the hormones that course through our veins and the guided missiles that target infections; they're the enzymes that build up and break down our energy reserves and the circuits that power movement and thought.*	VT	(+)
(102) lín. 18	*10,000 proteins*	EVint	*that's a tiny fraction of all the proteins found in nature*	VT	(N)
(103) lín. 37	*[The discovery of (basic) protein forms]*	EVint	[…] *robots and powerful X-ray generators have lately boosted the pace of discovery.*	VT	(+)
(104) lín. 46	*The flurry of private activity*	EVint	*raises the specter of intellectual-property disputes like those that plagued the Human Genome Project.*	AFE	(-)

10.2.2.3 Editoriales

Texto 10.3.1

Número/ línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría Valorativa	Valor (+) / (-) / (N)
(105) lín. 1	*This issue of 'Science'*	EVint	*appears amid a swirl of contemporary reminiscence about DNA* […]	AFE	(N)
(106) lín. 6	*[James Watson and Francis Crick's paper in which DNA structure was proposed]*	EVint	[…] *there was a splendid storehouse of information already available to provide that resonance.*	VT	(+)
(107) lín. 8	*[Undergraduates who believe that Watson and Crick's 'discovered' DNA]*	EVint	[…] *a claim those authors would be quick to disavow* […]	VT	(-)
(108) lín. 23	[…] *The transcription part of the problem*	EVint	*shifted* […] *to the knotty problem of how damage is repaired* […]	VT	(-)
(109) lín. 28	*Translation*	EVint	*became the key issue* […]	VT	(+)
(110) lín. 40	*[Establishing the relation between DNA sequences and the location of genes]*	EVint	*To do the same in eukaryotic cells proved a more difficult challenge.*	VT	(-)

Texto 10.3.2

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría valorativa	Valor (+) / (-) / (N)
(111) lín. 1	*The burgeoning commercial sector that is based on genome information*	EVint	*poses a challenge to the norms of scientific publication.*	JUI	(N)
(112) lín. 1	*[The human genome project]*	EVint	*The burgeoning commercial sector* […]	VT	(+)
(113) lín. 5	*[The human genome sequence]*	EVint	*It profoundly influences our bodies, our behaviour and our minds*	VT	(N)
(114) lín. 10	*[Investment in the Human Genome Project's determination of the sequence]*	EVint	[…] *it is hard to see how that investment could have received better returns.*	VT	(+)
(115) lín. 12	*'Nature'*	EVint	*is delighted this week to publish the project's analysis.*	AFE	(+)
(116) lín. 23	[…] *the human genome*	EVint	*is, in a symbolic sense, humanity's common heritage.*	JUI	(+)
(117) lín. 26	*Making freely available all the materials used for a piece of research*	EVint	*is sometimes impractical.*	VT	(-)
(118) lín. 32	*[Companies that make their living from data discovered through their own research]*	EVint	*There is big money at stake.*	VT	(N)
(119) lín. 54	*[Public and private scientists duplicating one another's work]*	EVext	*There are too many challenges around for it to make sense for public and private scientists to duplicate one another's work.*	VT	(N)
(120) lín. 56	*The private sector*	EVint	*is essential to the improvement of human health* […]	VT	(N)
(121) lín. 59	*The amount of valuable scientific information residing in research and development centres in the private sector*	EVint	*is huge* […]	VT	(+)

Texto 10.3.3

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría valorativa	Valor (+) / (-) / (N)
(122) lín. 1	*Good public health* […]	EVint	*rests on empirical data of high quality.*	VT	(N)
(123) lín. 14	*[Genetics looked unimportant compared to infectious diseases, high infant mortality and lack of proper sanitation]*	EVint	*With this shift, the burden of genetic disease has changed dramatically.*	AFE	(-)
(124) lín. 41	*("Personal contact*	EVint	*is the key to success").*	VT	(+)
(125) lín. 42	*[In the UK national health system the quality of the data depends upon the curator of the genetic register]*	EVint	*It is a pity that Professor Modell did not comment on this point* […]	AFE	(-)
(126) lín. 44	*[The UK national health system]*	EVint	[…] *it is encouraging that in the UK there were no problems about confidentiality* […]	AFE	(+)
(127) lín. 44	*[The UK national health system]*	EVint	[…] *families were confident that the register would be used in a positive and not in a discriminatory way.*	AFE	(+)
(128) lín. 54	*[The ethics of pregnancy choice]*	EVint	[…] *it is unethical for a country to push families towards screening* […]	JUI	(-)
(129) lín. 55	*[The ethics of pregnancy choice]*	EVint	[…] *it is equally unethical to deprive a couple of that information* […]	JUI	(-)

Texto 10.3.4

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría valorativa	Valor (+) / (-) / (N)
(130) lín. 3	*the newspapers*	EVint	*were full of headlines screaming of the benefits of the project to humanity.*	AFE	(+)
(131) lín. 37	*Brokering the peace between [a publicy funded consortium spearheaded by the Us National Institutes of Health and the Sanger Centre in Britain on one side and a private company, Celera Genomics on the other]*	EVint	*It is not inconceivable that this will happen again, only it will be for medicine.*	VT	(+)
(132) lín. 43	*[The availability of the genome sequences of diverse organisms]*	EVint	*From a purely scientific standpoint, biology research stands to benefit enormously from the availability of the genome sequences of diverse organisms.*	VT	(+)
(133) lín. 63	*A most distressing feature that has come to the fore*	EVint	*is the facility with which we tout our large (almost impossibly so) population and ethnic diversity as a major resource in biological research* […]	VT	(-)
(134) lín. 68	*[Achievements in genomics research]*	EVint	*Even in scientific circles, there is a worrying tendency to be self-congratulatory about our achievements in biological research* […]	VT	(-)
(135) lín. 77	[…] *our repeated drum beating on information technology (IT)*	EVint	*rests on the fragile fabric of a non-existent hardware base.*	VT	(-)

Texto 10.3.5

Número/ Línea	¿Qué es evaluado?	¿Quién evalúa?	¿Cómo se evalúa?	Categoría valorativa	Valor (+) / (-) / (N)
(136) lín. 4	*[The production of a first 'working draft' of the human genome]*	EVint	*[This] was a landmark in scientific research and a momentous occasion for 'Nature'* […]	VT	(+)
(137) lín. 15	*The collated information from these groups*	EVint	*would provide the most comprehensive analysis of the human genome.*	VT	(+)
(138) lín. 18	*[The work of the HGP]*	EVint	*This seminal work has* […] *been followed by the publication of the remaining chromosomes* […]	VT	(+)
(139) lín. 20	*['Nature']*	EVint	*In keeping with the spirit of the genomics community work ethic, 'Nature' has made each of these papers freely available on the Internet.*	JUI	(+)
(140) lín. 29	*[The publication of the completed chromosome analysis]*	EVint	*It was clear that they could not be coordinated in the same way.*	VT	(N)
(141) lín. 33	*[Each team had the time to complete their analysisto the highest level and to produce the best possible product without the work being lost in a competitive race to publish first, as is often the case in scientific research]*	EVint	*It was important that each team had the time to complete their analysis to the highest level and to produce the best possible product without the work being lost in a competitive race to publish first* […]	VT	(+)
(142) lín. 36	*'Nature'*	EVint	*fully appreciated the scientific value of publishing all 23 pairs of chromosomes as a single entity.*	AFE	(+)
(143) lín. 39	*[The publication by 'Nature' of the human chromosome papers as a single event]*	EVint	*The offer was for 'Nature' to review the work with the usual rigour bestowed on all of its scientific submissions* […]	VT	(+)