Efectos de la simplificación de textos en la comprensión lectora de las personas con discapacidad intelectual

Facultad de Psicología

Doctorado en Comprensión del Texto y del Discurso. Procesos Cognitivos y Aplicaciones Instruccionales

Valencia, 2015

Presentada por:

Gema Tavares Sanchez-Monge

Dirigida por:

Dra. Inmaculada Fajardo Bravo

Dra. Vicenta Ávila Clemente

La Dra. Inmaculada Fajardo Bravo y la Dra. Vicenta Ávila Clemente,

de la Universidad de Valencia,

DECLARAMOS

Que la tesis titulada *Efectos de la Simplificación de Textos en la Comprensión Lectora de las Personas con Discapacidad Intelectual* que presenta Gema Tavares Sanchez-Monge para la obtención del título de Doctora se ha realizado bajo nuestra dirección y cumple los requisitos para poder optar a la Mención Internacional.

Para que así conste y tenga los efectos oportunos firmamos el presente documento:

Dra. Inmaculada Fajardo Bravo

Dra. Vicenta Ávila Clemente


HAY MUCHAS FORMAS DE SER DISCAPACITADOS
LA ÚNICA PELIGROSA ES LA DE NO TENER CORAZÓN
©SENDRA

AGRADECIMIENTOS

Llegó el final de una etapa y, con él, el momento de reconocer y agradecer a quienes han caminado conmigo tanto tiempo, algunos aun sin saberlo.

Gracias al grupo Psicotext, hoy parte del ERI-Lectura, y especialmente a Inma y Vicen por lo muchísimo que me han enseñado en estos seis años.

Por supuesto, a Inma y Moi, por los buenos ratos y el apoyo y ánimo absolutamente incondicional. Con cada Bob Esponja os recuerdo con una sonrisa. A las compañeras de cafés y penas, Marian, Amelia, Pili y Mari. Los buenos recuerdos no entienden de distancias.

En un lugar muy especial quiero mencionar a Alberto, que tanto me enseñó y siempre creyó en mí (aunque lamentablemente las circunstancias me mandaron lejos); también a los profesores y compañeros de Salamanca, especialmente Laura y Carlos, por hacer que me picara el gusanillo de la investigación.

A AImée y Míriam, por vivir este proceso, a menudo no placentero, siempre a mi lado y disfrutando de los pequeños éxitos casi con la misma pasión que yo misma. A Carlitos y David, por entender sin juzgar y estar siempre dispuestos.

No hay palabras para agradecer a mi familia, mis padres, Víctor, Vito, Irene… por tantas y tantas horas de tratar de convencerme para no decaer, por pensar siempre que podía con todo y recordarme mis metas y el valor del esfuerzo aun cuando yo empezara a olvidarlo. Víctor, ¡cuánto lleva esta tesis de ti y qué poco has pedido a cambio!

El lugar más especial es para vosotros, Tita, Pepe y Jesús. Porque siempre hicisteis que quisiera ser mejor persona, más feliz y completa, y pensar que podéis estar orgullosos de mí me anima siempre a seguir avanzando. Os llevo conmigo. Siempre.

ÍNDICE

Glosario

GLOSARIO[1]

Adaptación/ Simplificación de textos

Proceso que tiene como objetivo la reducción de la complejidad lingüística de un texto manteniendo la información y el significado original (Siddharthan, 2014).

Lectura Fácil (LF)

Práctica similar a la simplificación del textos que conlleva también la adaptación ortotipográfica del mismo (fuente, disposición de los párrafos, etc.), y que es más empleada en el diseño de textos accesibles que en el campo experimental (Nomura, Nielsen, y Tronbacke, 2010).

Anáfora

(Uso de una) expresión que refiere a un elemento o concepto ya mencionado en el texto (Halliday y Hasan, 1976).

Conector

Elemento lingüístico que señala de manera explícita la relación entre dos proposiciones u oraciones consecutivas (Millis y Just, 1994).

Discapacidad Intelectual (DI)

Trastorno del neuro-desarrollo caracterizado por déficits en funcionamiento intelectual y adaptativo que se presenta antes de los 18 años (American Psychiatric Association, 2013).

[1] Las definiciones que incluyen una referencia bibliográfica son versiones simplificadas basadas en los trabajos y autores citados.

Familiaridad léxica

Característica cuantitativa de las palabras que indica el grado de incidencia de un término en la lengua según la valoración subjetiva de un grupo de personas (Crosson, Lesaux, y Martiniello, 2008).

Frecuencia léxica

Característica cuantitativa objetiva de las palabras que indica el número relativo de incidencias de un término en un corpus léxico (aquí, frecuencia por 2 millones. Ver Alameda y Cuetos, 1995).

Medida de legibilidad profunda

Cifra que representa la cantidad de elementos lingüísticos que sirven al mantenimiento de la cohesión de un texto (p.ej. número de correferencias) (Crossley, Louwerse, McCarthy, y Mcnamara, 2007).

Medida de legibilidad superficial

Cifra que representa la cantidad de elementos lingüísticos simples no relacionados directamente con la cohesión textual (p.ej. número de palabras).

1 Presentación de la tesis

1. Presentación de la tesis

Podríamos comparar la lectura con un mecanismo articulado por engranajes en el cual, de tanto en tanto, falla alguno de ellos sin que el funcionamiento global se vea afectado. Sin embargo, de vez en cuando el proceso se paraliza por ese mal funcionamiento individual. Por ejemplo, ¿cuántas veces hemos leído un texto y hemos llegado a comprender el mensaje pese a desconocer alguna palabra de las que contiene? Otras veces nos encontramos ante un texto cuyo significado global nos es indescifrable pese a que identificamos y conocemos todas las palabras que hay en él. El engranaje que falla no es siempre el mismo ni todos tienen la misma función. A estudiar parte de la complejidad de esta maquinaria dedicamos esta tesis, pero sirva aquí este símil para hacernos una idea aproximada de las dificultades con que pueden encontrarse las personas con un nivel de comprensión lectora limitado, como es el caso de los lectores con discapacidad intelectual (en adelante, DI) que constituyen el eje de este trabajo.

En particular, la presente tesis se presenta en la modalidad de "compendio de publicaciones" y está dedicada al análisis de las variables psicolingüísticas que afectan a la comprensión en la simplificación de textos dirigidos a lectores jóvenes de entre 16 y 23 años con DI. En este punto se hace necesaria la definición breve de dos conceptos que serán abordados con mayor profundidad a lo largo del compendio: 1) "simplificación de textos", que entendemos como el proceso de reducir la complejidad lingüística de un texto mientras se mantiene la información y el significado original (Siddharthan, 2014), y 2) Discapacidad Intelectual, entendida como un trastornos del neuro-desarrollo caracterizado por déficits en funcionamiento intelectual y adaptativo que se presenta antes de los 18 años (American Psychiatric Association, 2013). "Simplificación de textos", "adaptación de textos", diseño de textos de "Lectura Fácil" o "controlados lingüísticamente" son

todos términos similares[2] empleados para designar un conjunto de técnicas que con frecuencia se combinan con la instrucción directa para mejorar la comprensión lectora de los lectores con dificultades (Arfé, Oakhill, y Pianta, 2014; Lundberg y Reichenberg, 2013; Morgan y Moni, 2008; Quigley y King, 1981-1984). Como adelantábamos en el párrafo anterior, este sería el caso de los estudiantes con DI, un colectivo que usualmente presenta habilidades lectoras por debajo de la esperada para su edad (Channell, Loveall, y Conners, 2013; van den Bos, Nakken, Nicolay, y van Houten, 2007). Teniendo estos conceptos centrales definidos, pasamos a explicar la estructura de la tesis.

Dedicado pues a este tema, el compendio consta de 5 Estudios distribuidos en 3 artículos: cada artículo se compone de 2 Estudios, a excepción del primero, que contiene sólo uno.

Los artículos se han ordenado por fecha de realización y aceptación en sus respectivas revistas, no de publicación. Esto es debido a que la duración del proceso de publicación varió de una a otra revista o editorial por motivos ajenos a los autores. Así, el manuscrito del Artículo 1 fue enviado y aceptado para su publicación antes que el Artículo 2. Sin embargo, aun estando aceptado con anterioridad, el proceso editorial del Artículo 1 duró más de lo deseado y éste terminó siendo publicado con posterioridad. Con el fin de aclarar esta cuestión se incluye a continuación la Tabla 1 que presenta las fechas de aceptación y publicación de los 3 artículos.

[2] Al principio de la tesis se ha añadido un glosario de términos que, considerados fundamentales para seguir el hilo de la presentación, tienen, como éstos, significados similares o diversas acepciones según la disciplina o el área en el que se usen.

Tabla 1.

Fechas de aceptación y publicación de los artículos presentados.

Artículo	Nº de Estudio/s	Revista	FI	Aceptado	Publicado
1	1	JARID	1.098	Noviembre 2012	Mayo 2014
2	2.1, 2.2	RIDD	2.735	Enero 2013	Abril 2013
3	3.1, 3.2	RIDD	1.887	Diciembre 2014	Marzo 2015

Nota. FI = Factor de impacto en el área de Ciencias Sociales para el año de aceptación. JARID = Journal of Applied Research in Intellectual Disabilities (ISSN: 1468-3148); RIDD = Research in Developmental Disabilities (ISSN: 0891-4222).

A continuación pasamos a resumir los objetivos, metodología y principales resultados de cada Estudio, agrupados por artículos. En la sección *Anexos* se incluyen ejemplos de los materiales empleados, de manera complementaria a los contenidos en los propios artículos.

1.1. Artículo 1

En sus comienzos, la tesis fue concebida en el seno del proyecto "Lectura fácil: Elaboración y difusión de materiales y análisis de la experiencia de usuario"[3], que prosiguió bajo la denominación "Simplext"[4]. El grupo de investigación "ERI-Lectura" (Universidad de Valencia) del que la autora de esta tesis formaba parte estaba vinculado a dichos proyectos por medio de un contrato y un convenio respectivamente. El primer proyecto estaba liderado por Technosite, empresa dependiente de la Fundación ONCE, y tenía como objetivo la puesta en marcha de un periódico digital compuesto por textos simplificados o de "fácil lectura". Dicho

[3] Ministerio de Industria, Turismo y Comercio, 2008-2011 (TSI-040200-2008-30)
[4] Ministerio de Industria, Energía y Turismo, Plan Avanza 2, 2010-2013 (TSI-020302-2010-84)

periódico digital, que a día de hoy sigue en uso, quedó alojado en http://www.noticiasfacil.es/ y conformó la fuente de textos simplificados que se evaluaron en el Estudio 1.

El objetivo concreto del grupo "ERI-Lectura" era seleccionar y describir un grupo de lectores con DI a nivel de procesos de lectura, y comprobar empíricamente si los textos del periódico digital "noticias fácil" eran comprendidos por tales lectores. El Estudio 1 se llevó a cabo con el fin de abordar ese objetivo, por medio de un diseño descriptivo-correlacional con un único grupo experimental. Se incluyó también una amplia revisión de los antecedentes en la literatura sobre lectura y DI, así como en simplificación de textos, lo que nos permitió comprobar que el conocimiento previo del tema era escaso.

En este primer estudio se pidió a un grupo de participantes con DI que leyeran una selección de las noticias breves que Technosite había creado en Lectura Fácil (en adelante LF) siguiendo las pautas estándar publicadas por la Federación Internacional de Asociaciones de Bibliotecarios (IFLA) (Tronbacke, 1997), y respondieran a preguntas de comprensión. Con la finalidad de ir más allá del mero objetivo aplicado del proyecto, se trató de distinguir entre diferentes tipos de variables lingüísticas y hacer predicciones sobre su impacto a diferentes niveles de comprensión, todo ello guiado por el modelo de Construcción-Integración de Kinstch (1998). En este modelo la comprensión sucede a dos niveles: el lector toma la información explícita contenida en el texto (nivel literal) e integra esas ideas entre sí y/o con su conocimiento previo (nivel inferencial), construyendo así un Modelo de la Situación, una imagen completa de los actores y eventos descritos en el texto. En este mismo modelo se basan otros autores que trabajan en el ámbito de la simplificación de textos cuyos trabajos han servido de referencia en esta tesis, como por ejemplo Crossley y cols. (2007).

Así, de cada texto se obtuvieron diferentes índices lingüísticos que se agruparon, de acuerdo a la clasificación de los autores mencionados (op.cit., 2007), en medidas *superficiales* (p. ej. promedio de palabras por oración) y medidas *profundas* (p. ej. número de conectores discursivos). Asimismo se incluyeron dos tipos de preguntas de opción múltiple dirigidas a comprobar la comprensión tanto literal como inferencial del texto. En términos generales, se predijo una relación directa entre las medidas superficiales y la comprensión literal por un lado, y las medidas profundas y la comprensión inferencial por otro. Los resultados obtenidos revelaron que los textos adaptados daban lugar a niveles notables de comprensión por parte de los lectores con ID, tanto a nivel literal como inferencial, si bien en este segundo nivel la ejecución quedaba significativamente por debajo de la comprensión literal. Además, sólo algunas de las medidas exploradas resultaron estar relacionadas con la comprensión a uno u otro nivel, y a menudo las mencionadas predicciones de correlación no se cumplieron. En especial, este diseño descriptivo-correlacional nos permitió identificar que la cohesión textual no era un aspecto que quedara firmemente controlado aplicando las pautas estándar de simplificación aplicadas por los diseñadores de la noticias.

1.2. Artículo 2

A partir de los resultados obtenidos en el Estudio 1 diseñamos una serie experimental que ahondaría en la cohesión textual y la influencia de distintos elementos y recursos lingüísticos en la misma. Esta serie se compone de 4 Estudios distribuidos en dos bloques. El primer bloque (artículo 2), compuesto por dos Estudios (2.1 y 2.2), exploraría el efecto de los conectores discursivos sobre la cohesión textual atendiendo a su frecuencia (valor objetivo), familiaridad (valor subjetivo) y al tipo de relación semántica que éstos establecen (p.ej. aditiva,

consecutiva o temporal). A este bloque acompaña una revisión teórica que, consecuentemente, indaga en los efectos de la frecuencia léxica y los conectores en la comprensión lectora, tanto en niños con desarrollo típico como en adultos con dificultades de comprensión. Esta revisión nos permitió hacer predicciones basadas, por ejemplo, en la *Hipótesis de la Reactivación* de Millis y Just (1994) o en la *Teoría de la Complejidad Cognitiva* de Spooren y Sanders (2008).

El Estudio 2.1 consiste en un diseño cuasi experimental en el cual se puso a prueba la eficacia de la simplificación de textos a tres niveles diferentes: (1) adaptación de la frecuencia léxica de las palabras de contenido, (2) presencia/ ausencia de conectores, y (3) versión combinada de ambas pautas. Ni la presencia de conectores ni la versión combinada de conectores y frecuencia afectaron significativamente a la comprensión lectora de los participantes con DI. Dado que este resultado contradecía la experiencia previa de otros investigadores (Cain y Nash, 2011; K. Haberlandt, 1982; Koda, 2008; Millis y Just, 1994; Murray, 1995), en el Estudio 2.2 se manipularon dos parámetros lingüísticos que no habían sido controlados en el Estudio 2.1: la familiaridad y el tipo de relación semántica interclausal que establecen los conectores. Para los materiales se adaptó al español la tarea de comprensión de conectores tipo cloze de Crosson, Lesaux, y Martiniello (2008). En el grupo con DI observamos que la alta familiaridad facilitaba la comprensión cuando los conectores establecían relaciones semánticas aditivas, contrastivas y adversativas, pero la dificultaba en el caso de las temporales y causales. En la discusión del artículo 2 tratamos de explicar las consecuencias teóricas y aplicadas de esta interacción.

1.3. Artículo 3

El segundo Bloque, también conformado por dos Estudios (3.1 y 3.2), indagaría en el uso por parte de los lectores con DI de un elemento lingüístico diferente, relacionado también con la cohesión textual: las anáforas pronominales. La mayoría de los antecedentes que revisamos se centraban en hablantes nativos de inglés (Cain y Nash, 2011) o de francés (Ehrlich, Remond, y Tardieu, 1999), lenguas que, especialmente en el caso del inglés, poseen características y sistemas morfológicos diferentes al español. En otras palabras, no teníamos evidencia de cómo la manipulación de las anáforas pronominales podía afectar a la comprensión de un texto en español, ni mucho menos en el caso concreto de los lectores con DI, y los antecedentes en otras lenguas debían ser tomados con cautela. La revisión de la literatura realizada se centró esta vez en el procesamiento de las anáforas pronominales, así como en un elemento en auge en el estudio de la lectura, novedoso aún en la investigación en personas con DI: el registro de los movimientos oculares para la exploración de los procesos cognitivos involucrados en la lectura.

Este bloque comenzó con el diseño cuasi experimental del Estudio 3.1. En esta ocasión adaptamos la "prueba de anáforas" de Elosúa, Carriedo, y García-Madruga (2009). En dicha tarea los participantes deben leer una oración que contiene un pronombre clítico en posición pre-verbal con función de complemento directo (p.ej. Carmen *la* encontró en el aula al lado del pupitre) y a continuación escoger entre dos sustantivos cuál debería ser el antecedente del pronombre. La desambiguación del pronombre, es decir, la búsqueda del antecedente adecuado, debía realizarse bien mediante el uso de una estrategia morfosintáctica o de una estrategia semántica, para lo cual se contrabalancearon diferentes opciones de

respuesta (p.ej. *carpeta* vs. *cuaderno*, en el primer caso; *carpeta* vs. *profesora*, en el segundo caso).

De manera complementaria, el Estudio 3.2 proporcionaba un entorno más natural, pues los materiales eran textos, no oraciones aisladas, y los pronombres ya no aparecían en función de complemento directo sino de sujeto. El diseño de los materiales se basó en el paradigma de inconsistencias utilizado por Rinck, Gámez, Díaz, y De Vega (2003), que presenta textos de seis oraciones en los cuales la información temporal contenida en la última oración puede ser consistente o inconsistente con la información dada en la segunda oración. En nuestro caso la posible inconsistencia no era temporal sino morfosintáctica, y afectaba directamente a la desambiguación de un pronombre dado en la quinta oración, cuyos posibles antecedentes se encontraban en la segunda. El uso conjunto de textos completos y el registro de los movimientos oculares nos permitiría explorar en detalle cómo actúan los lectores con y sin DI antes y después de encontrar dicha ambigüedad. De manera complementaria, se añadió un conjunto de preguntas abiertas y de opción múltiple que se presentaron a continuación de cada texto. Estas preguntas apuntaban tanto a la comprensión literal como a la inferencial, así como a la detección explícita de la posible inconsistencia en el texto. Las preguntas de comprensión nos proporcionarían datos sobre la comprensión a posteriori (*offline*), mientras que el registro de los movimientos oculares arrojaría datos sobre los procesos activos durante la lectura (*online*).

2 Metodología y muestra

2. Metodología y muestra

Como ya se ha perfilado en el apartado anterior, los objetivos que fijamos para cada uno de los Estudios que componen este trabajo requerían de metodologías diferentes. El Estudio 1 se concibió como diseño descriptivo-correlacional por dos razones: 1) era un método que nos iba a permitir cumplir el objetivo del proyecto en el que se enmarcaba: ‘‘Lectura fácil: Elaboración y difusión de materiales y análisis de la experiencia de usuario” (esto es, comprobar el nivel de comprensión de los lectores con DI al enfrentarse a textos simplificados y tener una primera experiencia en la materia que nos ayudara a identificar posibles asociaciones entre las variables lingüísticas y las características del lector) y, 2) los materiales (textos) con los que teníamos que trabajar ya habían sido diseñados por otra de las entidades participantes en dicho proyecto, por lo que no teníamos opciones de manipulación sino simplemente de obtener medidas.

Tras las primeras experiencias con la muestra de lectores con DI en los Estudios 1 y 2.1, y tras revisar más a fondo la literatura sobre comprensión lectora decidimos que sería razonable incluir en las siguientes pruebas no sólo un grupo control de edad cronológica (jóvenes con un promedio de edad de 20.6 años), sino también uno de niños de nivel lector equivalente. La utilización de grupos de similares niveles en una habilidad concreta (en este caso la comprensión lectora) es un método frecuente en el estudio del lenguaje y la lectura (Cain, Oakhill, y Bryant, 2000). De esta manera puede identificarse si una habilidad particular (p.ej. las habilidades meta-cognitivas o fonológicas) es el resultado o no de un buen nivel de comprensión. Si el grupo experimental y el de similar edad de comprensión lectora difieren en habilidades meta-fonológicas, por ejemplo, será posible concluir que la primera no produce la segunda pero sí es una posible causa asociada al retraso o déficit de la comprensión lectora en el grupo experimental, y

debe ser explorada en mayor profundidad mediante diseños longitudinales o de entrenamiento. Además, este diseño permite responder al debate de déficit versus retraso en el desarrollo de la comprensión lectora que generalmente emerge en el área de los trastornos del desarrollo (ver, por ejemplo, Van der Molen, Van Luit, Jongmans, y Van der Molen, 2007). De acuerdo con la "hipótesis del retraso en el desarrollo", los lectores de bajo rendimiento con DI presentarían perfiles de comprensión lectora similares a los de los niños de desarrollo típico equiparados en edad lectora. Por el contrario, desde el punto de vista de la "hipótesis de la diferencia en el desarrollo", ambos grupos de lectores presentarían perfiles de comprensión lectora cualitativamente diferentes. Así, los Estudios 2.2 y 3.2 tienen un diseño cuasi experimental en el que participaron tanto un grupo experimental como dos grupos controles: uno de edad cronológica similar y otro de nivel lector equivalente. Como se indica más adelante en la sección “Limitaciones de la tesis”, por razones organizativas en los Estudios 2.1 y 3.1 únicamente se incluyó, junto al grupo experimental, un grupo control de similar edad cronológica.

Con respecto a las medidas de comprensión, se emplearon medidas *offline* y *online*, en particular el registro de movimientos oculares que, hasta donde sabemos, es novedoso en el ámbito de la lectura en personas con DI. Al buscar los términos "intellectual disability" (discapacidad intelectual) y "eye tracking" (movimientos oculares) en la base de datos ERIC (Education Resources Information Center, patrocinada por el Institute of Education Sciences), por ejemplo, obtenemos que desde 2006 únicamente se han publicado 7 artículos (20-6-2015), ninguno de ellos sobre lectura. A pesar de no haber sido suficientemente explotada aún, esta técnica tiene una importante ventaja en el ámbito de la comprensión lectora: permite la presentación normal del material de lectura, lo cual asegura que la adquisición de la información sucede de forma natural. Como

Rinck y cols. (2003) señalan, al poder visualizar el texto al completo los lectores tienen la libertad de inspeccionarlo, evitando los efectos no deseados que podrían aparecer en otras técnicas de presentación como el paradigma de la presentación "oración a oración".

Precisamente el haber llevado a cabo el registro de los movimientos oculares fue otro reto que resultó en una disminución del número de participantes (Estudio 3.2). Este método requiere que las pruebas sean administradas de forma individual, con condiciones ambientales muy controladas y por una persona que tenga conocimientos tanto de la investigación en desarrollo como de la parte técnico-informática para controlar el sistema de registro de movimientos oculares.

No sólo la metodología sino las características de la población experimental escogida han hecho el desarrollo de este trabajo especialmente complicado. Por un lado, el éxito de una prueba depende, en el caso de la investigación de los movimientos oculares, de ciertas características físicas de los participantes, pues condiciones como el estrabismo, algunos colores de iris, el uso de correctores de visión o incluso la incapacidad de permanecer sentado/a sin apenas moverse a lo largo de la prueba pueden hacer que los datos recogidos sean inservibles. Por otro lado, nuestros participantes eran jóvenes escolarizados en Camí Obert, un centro de formación profesional para personas con necesidades especiales que pertenece a APSA (Asociación Pro-Discapacitados Psíquicos), en Alicante. Al no tener acceso a las familias directamente, y dado que esta población suele padecer problemas médicos más o menos graves y situaciones personales complicadas, es frecuente que en estudios que requieren más de una sesión algunos participantes cursen baja antes de finalizar la totalidad de las pruebas. Además, debido a la heterogeneidad de las etiologías, solamente una pequeña parte de los alumnos del centro pudo completar las pruebas de referencia (CI, nivel lector, etc.).

Por estas razones la muestra presentada, especialmente el grupo de personas con DI, no es muy amplia. En el caso del grupo experimental, inicialmente se recogieron datos sobre todos los alumnos de Camí Obert que podían ser potenciales participantes. Las pruebas incluían datos de inteligencia verbal y no verbal (Kbit, Kaufman y Kaufman, 1997), nivel de vocabulario receptivo (Peabody, Dunn, Dunn, y Arribas, 2006), procesamiento sintáctico (subtest de la prueba PROLEC-SE, Ramos y Cuetos, 1999), memoria operativa (subtest de la batería WISC-IV, Wechsler, 2005) y comprensión lectora (ECL1, De la Cruz, 1999; TPC, Martínez, Vidal-Abarca, Sellés, y Gilabert, 2008). Con esta visión general y una vez aplicadas las condiciones especificadas en los párrafos anteriores, la muestra inicial de 50 estudiantes con DI quedó reducida a 12-28 (según el Estudio), algunos de los cuales participaron en varios Estudios. La Tabla 2 muestra los datos de participación por grupo y Estudio.

Tabla 2.

Distribución de los grupos a través de los 5 Estudios.

	GE	GC-EC	GC-EL
E1	16 (28)	-	-
E2.1	19 (28)	-	-
E2.2	16 (17)	18 (28)	18 (51)
E3.1	28 (29)	27 (27)	-
E3.2	12 (19)	16 (19)	12 (25)

Nota. Tamaño de la muestra final y tamaño muestra inicial (entre paréntesis). GE = grupo experimental, GC-EC = grupo control de similar edad cronológica, GC-EL = grupo control de edad lectora equivalente.

En el caso del grupo experimental se fijaron dos criterios de inclusión:

1. Cociente Intelectual (CI) entre 40 y 85, rango que según el manual de diagnóstico DSM-IV-TR (American Psychiatric Association, 2000)

corresponde a un nivel de severidad de límite a moderado[5]. Dado que al principio de esta tesis la 4ª edición del manual DSM era la vigente, por razones de consistencia se ha empleado esta versión a lo largo de todo el trabajo. El nivel de gravedad de DI se ha establecido, por lo tanto, en función del CI, y no como propone el DSM-V (American Psychiatric Association, 2013) en función de la conducta adaptativa.

2. Nivel de lectura mínimo equivalente a 2º de Educación Primaria (EP). En el currículo español de Educación Primaria se entiende que al completar el 2º curso el estudiante debe haber adquirido unas competencias mínimas en cuanto a la comprensión de textos escritos. Estas incluyen, literalmente, la "Comprensión de textos de la vida cotidiana, próximos a la experiencia infantil […] y de los medios de comunicación (noticias) […]." (Decreto 111, 2007). Hemos considerado este el nivel mínimo imprescindible para que los participantes pudieran enfrentarse al tipo de textos que se iba a evaluar.

Como único criterio de exclusión se estableció el tener un co-diagnóstico de Trastorno del Espectro Autista (TEA), aspecto que verificó el centro educativo "Camí Obert" mediante los expedientes escolares de los estudiantes.

En el caso de los grupos control únicamente se controlaron la comprensión lectora y el CI en el grupo de nivel lector equivalente para asegurar que cumplía con el nivel de 2º EP. Puesto que el grupo control de similar edad cronológica estaba conformado por estudiantes universitarios, se asumió que ambos parámetros estaría dentro de la normalidad y no se realizaron más pruebas aparte de las experimentales.

[5] A pesar de que la DI límite no cumple el criterio de estar dos desviaciones típicas por debajo de la media (ca. 70), razón por la cual se considera "categoría adicional" en el DSM-IV-TR, a lo largo de este trabajo sí se ha considerado dentro de la escala de DI en pro de una simplicidad lingüística de las definiciones.

3 Study 1

Easy-to-read Texts for Students with Intellectual Disability: Linguistic Factors Affecting Comprehension.

Fajardo, I., Ávila, V., Ferrer, A., Tavares, G., Gómez, M., & Hernández, A. (2014). Easy-to-read texts for students with intellectual disability: linguistic factors affecting comprehension. *Journal of Applied Research in Intellectual Disabilities: JARID*, 27(3), 212–25. http://doi.org/10.1111/jar.12065

Easy-to-read Texts for Students with Intellectual Disability:

Linguistic Factors Affecting Comprehension

Inmaculada Fajardo*, Vicenta Ávila*, Antonio Ferrer*, Gema Tavares*, Marcos Gómez-Puerta†, Ana Hernández‡

*ERI Lectura, University of Valencia; † Department of General Didactics and Specific Didactics, University of Alicante (Spain); ‡ CFP Camí Obert, Alicante

Abstract

Background. The use of 'easy-to-read' materials for people with intellectual disabilities has become very widespread but their effectiveness has scarcely been evaluated. In this study, the framework provided by Kintsch's Construction–Integration Model (1988) is used to examine (i) the reading comprehension levels of different passages of the Spanish text that have been designed following easy-to-read guidelines and (ii) the relationships between reading comprehension (literal and inferential) and various linguistic features of these texts.

Method. Sixteen students with mild intellectual disability and low levels of reading skills were asked to read easy-to-read texts and then complete a reading comprehension test. The corpus of texts was composed of a set of forty-eight pieces of news selected from www.noticiasfacil.es, a Spanish digital newspaper that publishes daily journalistic texts following international guidelines for the design of easy-to-read documents (IFLA, Tronbacke B. (1997) Guidelines for Easy-to-Read Materials. IFLA, The Hague).

Results. Participants correctly answered 80% of the comprehension questions, showing significantly higher scores for literal questions than for inferential questions. The analyses of the texts' linguistic features revealed that the number of coreferences was the variable that best predicted literal comprehension, but contrary to what the previous literature seemed to indicate, the relationship between the two variables was inverse. In the case of inferential comprehension, the number of sentences was a significant negative predictor; that is, the higher the sentence density, the lower the ability of these students to find relationships between them. The effects of the rest of linguistic variables, such as word frequency and word length, on comprehension were null.

Conclusions. These results provide preliminary empirical support for the use of easy-to-read texts but bring into question the validity of some popular design guidelines (e.g. augmenting word frequency) to optimally match texts and reading levels of students with intellectual disability. Two factors are suggested as contributing to the effect of sentence density on inferential compre- hension: (i) long texts present higher conceptual density, so there are more ideas to store, retrieve and integrate, which increases the demand on inferential reasoning and (ii) long texts are perceived as difficult, which affects reading motivation and, consequently, induces passive reading strategies. The need for further research to elucidate the origin of our main findings with a larger and more heterogeneous sample of students with intellectual disability is highlighted.

Keywords: easy-to-read texts, linguistics, readability measures, reading comprehension.

Easy-to-read Texts for Students with Intellectual Disability:

Linguistic Factors Affecting Comprehension

Literature review

Reading in general and text comprehension in particular have been identified as challenging skills for students with intellectual disability (ID), who usually do not achieve the level of their chronological and educational peers (Conners, 2003). The use of "easy-to-read", or simplified, texts is one of the strategies used by educators to improve the reading comprehension performance of these types of students. However, the design or selection of "easy-to-read" texts is not a straightforward issue, so educators usually use recommendations made by international organizations such as the *Guidelines for Easy-to-Read Materials* by the International Federation of Library Associations and Institutions (IFLA) (Tronbacke, 1997) or the *Make it simple Guidelines* by the International League of Societies for Persons with Mental Handicap (ILSMH) (Freyhoff et al., 1998).

However, as guidelines designers admit (Nomura, Nielsen, & Tronbacke, 2010), further research is needed in order to provide empirical support for the mentioned sets of guidelines, especially in the case of readers with ID. In addition, the extent to which a text fulfils a particular guideline or set of guidelines is not easily quantifiable. For instance, one example of the *Make it Simple Guidelines* set is to "avoid difficult words". How could educators objectively measure and compare texts that differ in this variable? The application of readability formulae based on text linguistic measures such as word frequency and word length (e.g. the classic Flesch's formula by Flesch, 1948) would aid in this process but, to the best of our knowledge, they have scarcely been tested empirically for the purpose of selecting

easy-to-read texts for readers with ID. Finally, easy-to-read guidelines designers and many readability formula developers do not explicitly state the theoretical models of reading and reading comprehension framing their work, if any, so it is difficult to establish their construct validity. Therefore, using the framework provided by Kintsch's (1988) construction-integration model, the aim of this study was twofold: (1) To test intellectually disabled students' reading comprehension performance on a corpus of texts designed according to the *Guidelines for Easy-to-Read Materials* (Tronbacke, 1997) and (2) To examine the relationship between texts' linguistics variables, on the one hand, and reading comprehension performance at both literal and inferential level, on the other, in order to better establish how to measure text readability.

The following section presents a brief review of the reading process in intellectual disability, which will help to frame the objectives of our research. That section is followed by a description of classical and new linguistic measures applied in the fields of readability and easy-to-read text design. Finally, the method and results of our study are explained and interpreted.

Reading in students with intellectual disability.

A general language deficit or developmental delay is commonly identified as the core process affected in individuals with intellectual disability, especially in some aetiologies like the Down syndrome (Rondal, 2001). This general language problem might be influencing the acquisition of reading literacy although, paradoxically, some reading processes seem to be acquired in spite of such a deficit. For instance, individuals with Down syndrome are able to recognize words without good phonological awareness considered a pre-requisite of reading for

regular readers (Fowler, Doherty & Boynton, 1995). However, as Fowler et al. suggest, word recognition and listening comprehension does not guarantee reading comprehension.

Therefore, this type of findings seems to highlight that reading is a process which composed of multiple components each of which should be considered in order to adequately measure individuals' competence. In fact, defined as the process of extracting meaning from print, reading is far from being a simple process and can be roughly divided into two components: (1) lexical decoding (the process of transforming printed word into a sound) and (2) comprehension (the process of understanding the literal and implicit ideas of a text). With regard to the former component, previous research has shown that most children with ID show weak lexical decoding skills (e.g. Conners, Atwell, Rosenquist, & Sligh, 2001) and a small mental lexicon, knowledge of words (Aitchison, 1994) or reduced vocabulary (e.g. Ferrer et al. [1999] observed that people with ID aged 14 presented an equivalent age of 4-5 years in receptive vocabulary). Although strongly related, the focus of the present work was not decoding but comprehension, which according to Kintsch's Construction-Integration Model (1988), occurs at two levels: (1) literal level: comprehension of the actual meaning of single propositions and (2) inferential level: integration between text segments or between text and prior knowledge. These two sentences may serve to illustrate the extraction of literal and inferential meanings:

The politician said that petrol prices would decrease by 10% in 2010. She was wrong.

If a teacher asked his/her students after reading these sentences "How much would the prices decrease in 2010 according to the politician?", the correct answer would be 10%; a piece of information which is literally stated in sentence

1. If the question was “Did petrol prices decrease by 10% in 2010?” the correct answer would be “No, prices did not decrease a by 10%”. This piece of information must be inferred from the text by integrating ideas from sentence 1 and sentence 2 (the politician was wrong, so prices did not decrease by 10% in 2010). In addition, a student could use their prior knowledge about petrol price changes in 2010 in order to answer the second question (e.g. his/her father drove less in 2010 so petrol prices probably increased, rather than decreased). The first and second answers constitute what Kintsch calls the textbase model and it is based on the “construction” of a mental representation of the text. The third answer requires the “integration” of the reader’s knowledge and experience with the mental representation of the text and is part of the reader’s situation model.

The question that arises here is ‘what is known about these two comprehension processes (literal and inferential) in intellectual disability?’

Literal and Inferential levels of reading comprehension. Regarding literal levels of comprehension, the low levels of vocabulary knowledge mentioned above suggests that the actual meaning of single sentences may not be understood if they are composed of low frequency content words that are not in the lexical repertory of students with ID. Besides, these students may present problems with functional words such as pronouns and prepositions, which are also essential syntactic cues that help us to understand clauses. Perovic (Perovic, 2006) found that young adults with Down syndrome and mild intellectual disability presented difficulties interpreting reflexives (e.g. herself, himself) but not pronouns (e.g. him, her) in an oral comprehension task. The author concluded that, as this is the opposite pattern of acquisition to that found in typically developed children (first reflexives, then pronouns), language acquisition in Down syndrome is not just

delayed but qualitatively different at the specific level of the morphosyntactic processing.

Although the delay versus deficit debate is a research question expandable to intellectual disability in general, reading comprehension in Down syndrome presents particular features (e.g. specific deficits in phonological awareness and short term memory, see Fowler, Doherty & Boyton for a review) not presented in others intellectual disability etiologies. Due to this particularity, we did not include students with Down syndrome in the present study.

Although the understanding of pronouns like reflexives can influence the literal comprehension of single sentences, it can also affect the inferential level of comprehension since they usually work as linking devices, connecting parts of the text that need to be integrated in order to get implicit ideas. The ability of children with ID to infer pronouns and other types of linking devices, like connectives (e.g. however, in addition, etc.), was measured by Verhoeven & Verneer (2006) with a cloze task. Children with ID were asked to complete text fragments in which those elements had been omitted. Their scores in this task were significantly lower than those of typically developed children. A possible interpretation of these findings is that readers with ID find it difficult to understand the type of relationship established between sentences (e.g. causality or temporality) and consequently they cannot correctly guess the connective that should accompany them. Sanders et al. (2007) hypothesise that connectives could actually help poor readers to integrate sentences. Effectively, they found that individuals with poor comprehension levels (non-ID), obtained higher comprehension scores after reading texts with connectives than the texts without them, so we could predict a similar facilitative effect of connectives in students with ID.

In order to measure literal and inferential comprehension of narrative texts, Wolman (1991) asked children with mild ID to recall and judge the relevance of recalled statements of cohesive and non-cohesive versions of two stories. Children who were able to make inferences would understand the meaning of the story, differentiating relevant from irrelevant information, as shown in their judgements of relevance. Causal cohesion of stories was manipulated by varying the proportion of causal chain statements in the text; the higher the proportion of causal chain statements, the higher the casual cohesion. The group of children with ID recalled significantly less statements and episodes of the stories, mainly in the non-cohesive stories, than children with learning disabilities (children with reading difficulties and normal intelligence) and children without disability (children with normative levels of reading and intelligence). This finding demonstrates that literal comprehension of the stories was lower in children with ID than in the other two groups. In addition, children with ID judged relevant and irrelevant statements as equally important, especially in non-cohesive stories; i.e. children with ID find it difficult to infer the meaning of the story, which would help them to discriminate relevant from relevant information.

Finally, regarding specifically inferential comprehension in students with ID, Bos & Tierney (1984), asked children with and without ID to read narrative and expository texts and free-recall the text content. Narrative texts tell a story and have beginning, middle and end, characters, plot or conflict, and setting. Expository texts inform or describe and the information is organized in a logical manner. The analysis of free-recall revealed that the number and quality of inferences were identical for both groups in the narrative text condition. In the case of expository text, children without ID outperformed children with ID in the quality of inferences. The authors suggest that the narrative scripts are more accessible than expository

scripts for children with ID, which would help them to make reasonable inferences about events in narrative texts. So far, the insight obtained from this finding is that children with ID are able to make inferences if the structure of the text is well known to them, as in the case of narrations.

In addition to specific linguistic and cognitive factors, there are subjective and motivational factors involved in reading comprehension as well. As detected in previous studies with ID youth readers, they usually do not read autonomously; feel frustrated or are reluctant to read long texts and texts with technical difficulty (e.g. academic texts). However, their attitude is highly receptive to the use of Internet reading environments, for example, digital newspapers (Gómez-Puerta, 2011; Morgan & Moni, 2008) where in addition to multimedia content, they can find personally interesting texts. That was the type of texts used in the present study.

This brief review of the topic confirms that children with ID present problems at both levels of Kintsch's Model of reading comprehension; i.e. literal and inferential. In particular, they can find it difficult to: understand the meaning of single sentences because of a lexical deficit; recall main statements of the text and identify main ideas; and interpret functional words and other syntactical cues that help to make inferences. We have also learned that findings from research in typically developed children must not be generalized for children with ID because their pattern of reading behaviour can be different, not just delayed (Perovic, 2006).

Now that the profile of readers with ID has been outlined, how can we know if a text fulfils the lexical, syntactical and cohesion constraints required for them? The following section presents a description of the most popular psycholinguistic measures in the fields of legibility and easy-to-read text design.

Easy to read text design guidelines and psycholinguistic measures.

As we mentioned above, there are sets of guidelines for the design of easy-to-read texts produced by several international organizations (e.g. IFLA and ILSMH Guidelines) but there is little empirical research that supports them, especially in the case of readers with ID.

Karreman, Van Der Geest and Buursink (2007) have conducted one of the few empirical studies testing the comprehension levels of individuals with ID after reading two different versions of a website: one adapted on the basis of ILSMH's easy-to-read guidelines (1998) and the other non-adapted. They found that both literal and inferential comprehension of the group of individuals with ID was higher in the adapted version than in the non-adapted version of the website. Although this finding is very useful, the drawback of the methodology used in this research is that the adapted and non-adapted websites differed in a number of linguistic elements (e.g. length of words and sentences, frequency and abstractness of words, tense of sentences, etc.), which makes it difficult to distinguish which elements or collection of elements were in effect facilitating reading comprehension.

In addition to the indicated lack of exhaustive examination of the validity of specific IFLA or ILSMH guidelines, the problem with some guidelines is that they are formulated in very general terms, making it difficult to use them. For instance, two of the IFLA guidelines are "use a simple, straightforward language" and "use a clear and logical structure". How can text designers implement or measure these types of features? Readability or legibility formulae may serve to aid in the measuring process if a text fulfils or does not fulfil some of them.

As suggested by Crossley, Dufty, McCarthy and McNamara (2007), we can talk about shallow and deep readability formulae. In the former group, the measurement of text difficulty relies on superficial text variables, such as word and sentence length (e.g. Flesch Index by Flesch, 1948), or word frequency (CELEX by Baayen et al., 1995). It is assumed that the shorter the word and sentence length and the higher the word frequency, the lower is the text difficulty. However, there is contradictory evidence concerning this. While the classical study of Doctorow, Wittrock and Marks (1974) showed that children's reading comprehension was increased when text with high frequency words was provided, other authors have found a null effect of increasing word frequency on the literal and inferential comprehension of texts (Freebody & Anderson, 1983; Ryder & Hughes, 1985).

The effect of word frequency is not clear either in the case of simpler lexical tasks, such as naming or lexical decisions. Although it has been classically reported that frequent words are read more quickly than infrequent words (Frederiksen & Kroll, 1976), null or even inverse effects of frequency have also been found (Adelman, Brown, & Quesada, 2006; Monsalve & Cuetos, 2001).

In addition to the lack of clear empirical evidence about the effect of shallow readability measures, these types of formulae are insufficient for evaluating other text features, such as cohesion. In the attempt to overcome the problems with traditional readability formulae, the Coh-Metrix computational tool (Graesser, McNamara, Louwerse, & Cai, 2004) measures cohesion features, such as the density of connectives and co-references, as it is assumed that those elements would help readers to link ideas and construct a coherent mental representation of the text content. For example, Crossley et al. (2007), measured the number of connectives in authentic texts (text written to fulfil a social purpose for native

speakers within a language community, Lee, 1995) and simplified texts for second language students. They found that simplified texts tend to contain more common connectives (e.g. and, but) and more complex and infrequent connectives (e.g. if, only if, on the condition that, etc.) than authentic texts.

However, Crossley et al. (2007) noted that the syntactic simplification may not necessarily be facilitating text comprehension because such a type of simplification could prevent readers from elaborating ideas that are needed to understand hypothetical situations. For instance, let's consider the following paragraph written by a journalist: "If politicians were sensible they would search out the economists that predicted the crisis". The hypothetical situation supposed by the journalist would be more difficult to infer in a paragraph like this "Politicians are not sensible and they did not search out the economists that predicted the crisis", where the syntactic complexity is lower than in the "authentic" text but the journalist's hypothesis about the relationship between politicians sensibility and the lack of prevention of the crisis is not evident.

Another cohesion element is the co-reference, defined as argument overlaps between sentences, which seems to improve text comprehension and reading speed (Kintsch & Van Dijk, 1978). For instance, in the following two sentences "Anne ordered Claudia to sit down. She is very bossy", "she" works as a "pronominal co-reference" for "Anne". If the proper name "Anne" was repeated in the second sentence instead of "she", then it would be working as a "same noun phrase repetition" co-reference. McNamara, Louwerse, McCarthy and Graesser (2010) assessed the validity of a set of text cohesion indices of Coh-Metrix. The researchers checked the ability of these indices to discriminate between high and low-cohesion versions of texts. Texts were extracted from published experimental studies in which the variable "cohesion" had been manipulated between groups,

that is, where different groups of participants had been asked to read more cohesive and less cohesive versions of the same texts. McNamara et al. found that, in effect, more cohesive texts contained more causal connectives and co-references (in particular, noun, arguments and stem overlaps) than less cohesive texts. The validation of these deep-level linguistic measures would suggest that they must be taken into account in order to design or adapt texts for students with ID.

Finally, it is important to note that the text adaptation approach for students with ID is frequently aided with the used of pictures, icons and symbols despite the fact that the benefits of them are not well established yet. As a matter of fact, Jones, Long and Finlay (2007), observed that the addition of symbols to texts improved reading comprehension of adults with mild and borderline ID. On the other hand, the study of Poncelas and Murphy (2007) compared two versions of a simplified text: one text-based and the other symbol-based (with text). Adults with ID were asked a series of questions about the material to assess comprehension. Both versions produced low levels of comprehension. The group of readers with symbols (and text) showed no better understanding than the group with text only. The results of these two studies would suggest that the addition of symbols to simple texts does not necessarily improve readers' comprehension of it.

It is important to note that methodological differences between both studies (e.g. types of symbols, type of texts, age mean of the sample, etc.) may account for the difference in effectiveness of addition of symbols to texts. Therefore, as the former authors suggest, future research should focus on the control of those methodological parameters in order to elucidate the origin of symbols reading comprehension facilitation. While admitting the importance of graphic aids, the

present study focused exclusively in the isolation of linguistics factors affecting literal and inferential reading comprehension for readers with ID.

This literature review makes it evident that the recommendations for the design or selection of easy-to-read texts are neither proven nor easily quantifiable for students with ID. It seems necessary to identify which linguistic features must be considered to make a text easy to read for students with ID. Therefore, we designed a descriptive study aimed at: (1) Testing the reading comprehension levels of a corpus of easy-to-read texts by intellectually disabled students, (2) Examining the relationship between texts' surface and deep linguistic measures, on the one hand, and literal and inferential reading comprehension performance on the other.

In particular, our research questions were: (1) What is the reading comprehension level of a corpus of easy-to-read texts by intellectually disabled students? (2) Do surface and deep linguistic measures predict literal and inferential reading comprehension in students with ID?

Method

Participants

Twenty-eight students with ID (17 males and 11 females) were recruited in 'Camí Obert', a centre of vocational training belonging to the Asociación Pro-Discapacitados Psíquicos of Alicante. Students and students' parents signed an informed consent before participating in the study. The average age of the sample was 18.9 (*SD*=1.84, ranging from 16 to 22). No individuals with known genetic syndromes, in particular Williams' and Down's syndromes, were included in the sample. The reason for this exclusion criterion was that those specific syndromes

present language particularities which could be not present in the rest of aetiologies and they did not constitute a big enough group as to be studied as a separated group.

In order to establish the baseline of reading skills, the participants were asked to complete a series of reading processes and reading-related linguistic tests: Spanish Peabody Picture Vocabulary Test (PPVT-III) (Dunn, Dunn, & Arribas, 2006), syntactic processing subtest from the PROLEC-SE test (Ramos & Cuetos, 1999) and the ECL1 Reading Comprehension Test (De la Cruz, 1999). The PPVT-III measures the individual's receptive vocabulary by asking participants to identify words based on a series of pictures. The examiner states a word describing 1 out of 4 pictures and asks the individual to point to or say the number of the picture that the word describes. The syntactic subtest of the PROLEC-SE consists of a set of pictures that represents scenes (for instance, a policeman arresting a burglar). Participants are asked to select one written sentence out of four that correctly describes the scene. The syntactic complexity of the sentences increases as trials proceed so that the syntactic competence of participants can be established. The ECL1 measures reading comprehension in children aged 6 to 8 years and consists of 3 short texts (approx. average sentence length of 100 words) accompanied by literal and inferential questions. In addition, the participants' IQ was measured with the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1997).

Twelve participants were removed from the original sample because they did not fulfil the criterion of mild to moderate mental retardation according to the DSM-IV-TR Manual (American Psychiatric Association, 2000). The average IQ score measured with the K-Bit test for the remaining 16 participants was 64.9 (*SD*=10, min=43, max=82). Corrected scores for each measure were obtained

from the normative data of the higher, either educational level, or chronological age provided by each particular test. The remaining 16 participants fulfilled the criterion of having a reading comprehension level of at least second year primary school, measured with the ECL test, although only 56.2% of the sample reached the level of third year primary school (raw score, *M*=9.1; *SD*=2.4). For the rest of the reading tests, the participants performed below their chronological age (Peabody´s raw score, *M*=110.9; *SD*=22; Equivalent age= 9.3) or educational level (PROLEC-SE's raw score, *M*=12.9; *SD*=4.3; percentile of fourth-grade secondary school= 5).

Task and materials

Participants were asked to read easy-to-read texts and then complete a reading comprehension test. This study was framed by a project aimed to increase the usage, usability and validity of a Spanish digital newspaper, www.noticiasfacil.es, which publishes daily journalistic texts. These news or journalistic texts are directly designed (written) in an easy-to-read format following the IFLA's guidelines. The newspaper uses this particular set of guidelines because they form the standards for easy-to-read text adaptation and design more commonly found internationally and the only one adapted to Spanish. Therefore, the corpus of easy-to-read texts was composed of a set of 48 pieces of news selected from this website. The genre of texts was narrative or expository. There were 6 to 10 pieces of news for each of the newspaper's six main sections: economy, sport, national, international, culture and miscellaneous. For each piece of news, three levels of linguistic variables were measured: lexical and sentence variables (shallow measures), and textual cohesion variables (deep measures).

Lexical variables. Number of words, average syllables per word and average word frequency. Word frequency was obtained from the Dictionary of Spanish Linguistic Units (Alameda & Cuetos, 1995), a Spanish data base of two million words in which word frequency is measured as the number of occurrences per million words.

Sentence variables. Number of sentences per text, average words per sentence and Flesch-Szigriszt Index (Szigriszt-Pazos, 1992), a version of the classical Flesch index for texts in Spanish. The Flesch-Szigriszt Index is expressed as follows: 206,835 - (62.3 x S/P) - P/F, where S is the number of syllables, P is the number of words in the text and F is the number of phrases (defined as the content between two punctuation marks).

Textual cohesion variables. Referring to elements affecting the integration between sentences and paragraphs. In particular, in this research, we measured: (1) the number of co-references – noun overlap (where the same noun is common to a pair of sentences; e.g. child-child), argument overlap (where the word stem is repeated among sentences without varying the lexical category; e.g. child-children) and stem overlap (where the shared form is just the word stem, allowing word category variation; e.g. child-childish); (2) the number of ellipses (omissions in a clause of one or more required words or elements that are implicit and can be understood anyway); and (3) the number of connectives per text (linking words, such as conjunctions between ideas or clauses; e.g. however, even though, after that).

The reading comprehension test was composed of multiple-choice questions that tested readers' comprehension of the news content. These questions were compiled according to the two levels of processing (literal and inferential) of the Kintsch's Construction-Integration Model (1988). Literal

questions: the answer to this type of question was either explicitly stated or restructured within a single sentence. Inferential questions: required the integration of inter-sentence information and integration of textual information with prior knowledge. The number of literal and inferential questions varied per text because it was not always possible to extract them from the information provided for the texts (see examples in Appendix 1). In particular, for 3 pieces of news there were no inferential questions so at this level only 45 pieces of news were analysed. One piece of news did not have any literal questions, so the analysis of this type of question was performed for 47 texts only. The total number of questions was 178 (100 literal questions and 78 inferential questions).

Multiple choice questions were used for several reasons: 1) To prevent the potential masking effects of writing competence and memory or verbalization difficulties of our students which could appear with open questions such as retelling or another type of free recall technique (for a discussion about this issue see Fletcher & Clayton, 1994). In order to prevent the additional difficulty of the questions themselves, both questions and choices were short and written in easy-to-read format; 2) To facilitate answer analysis and interpretation due to the high number of questions we had (178 in total); 3) To ensure that readers had to activate a specific type of process in order to correctly answer our questions since we were interested in measuring two different types of reading processes: literal vs. inferential. The correct choice for multiple choice questions can be selected by researchers in such a way that readers necessarily had to activate one of the referred processes to find it. However, the process activated by open questions could be more dependent on the reader's question answering satisfaction threshold than on the information comprehension itself.

Procedure

Each week, for a period of 16 weeks, participants were asked to read three pieces of news selected randomly from www.noticiasfacil.es and to complete a reading comprehension test. Thus, there were 48 pieces of news in total. Each week, participants were given a booklet with the three texts, each followed by comprehension questions on it. The order of presentation of the texts was the same for every participant since, as pieces of news, the relevance of the topic for each participant could have varied if the presentation order had been varied throughout the days. The order of presentation of literal and inferential question options was randomized for each text. Although participants were instructed to first read each text and then answer the questions, they were allowed to go back to the text if necessary. When the questions for a text had been answered, the participants were instructed to read the next text. There was no time limit for completing the task. Text reading and question fulfilment processes were supervised by the participants' educators ensuring that their doubts about the question answering process were solved before starting.

Results

What is the reading comprehension level of a corpus of easy-to-read texts by intellectually disabled students?

Our first goal was to verify if easy-to-read news texts were, in effect, comprehended by students with ID. Therefore, an ANOVA test was performed for each item (piece of news), with the percentage of correct answers for literal and inferential questions used as measures (N was 47 and 45 respectively). The percentage of correct answers to literal and inferential questions was .87 (*SD*=.12)

and .76 (*SD*=.22) respectively, with the difference between two types of questions being significant, *F* (1, 43) =11, 26; *MSE* =.248; *p* <.0017.

Do surface and deep linguistic measures predict literal and inferential reading comprehension in students with ID?

Correlation analyses. In order to address the second goal of the study (i.e. to examine the relationship between the texts' surface and deep linguistic measures and reading comprehension performance), several Pearson correlation analyses were performed between each block of linguistic measures on the one hand and comprehension measures on the other. The descriptive data for each linguistic variable is provided in Table 1.

With regard to lexical variables, we found significant negative correlations between the number of words and both literal and inferential comprehension questions; i.e. the longer the texts, the lower the literal and inferential comprehension of the students. Neither word length (number of syllables per word) nor word frequency correlated significantly with comprehension scores.

With regard to sentence variables, we found significant negative correlations between the number of sentences and both literal and inferential comprehension; i.e. the higher the number of sentences, the lower the literal and inferential comprehension of the texts. However, no significant correlations were found between comprehension scores on the one hand and average words per sentence and Flesch-Szigriszt index on the other.

Finally, with regard to textual cohesion variables, we found significant negative correlations between the number of connectives and literal but not inferential comprehension. The total number of co-references correlated

significantly and negatively with literal comprehension; i.e. the higher the number of co-references contained by texts, the lower the readers' scores for literal questions. This correlation was not replicated in the case of inferential comprehension questions. Similarly, the number of ellipses did not correlate significantly with comprehension scores.

Table 1.

The first and second columns show Ms and SDs for each level of linguistic measures of the text corpus (N48) selected for this study. The third and fourth columns show Pearson correlation coefficients between linguistic measures and literal and inferential comprehension questions.

			Percentage of correct answers M (*SD*, *Mdn*)	
Lexical variables	M	SD	Literal questions (47 texts)	Inferential questions (45 texts)
No. of Words	93.8	40.9	-0.30*	-0.32 *
Average syllables per word	1.9	0.1	-0.14	0.05
Average word frequency	21361	4427	-0.16	0.05
Sentence variables				
No. of sentences	11.6	5.6	-0.30 *	-0.37 *
Average words per sentence	8.3	1.5	-0.03	0.11
INFLESZ index	74.7	6.7	0.08	-0.11
Textual cohesion variables				
No. of connectives	4.0	3.3	-0.35 *	-0.13
No. co-references (noun, argument and stem overlap)	12.5	6.2	-0.39 **	-0.24
No. of ellipsis	2	2	-0.22	-0.24

Marked values are significant at *$P < 0.05$ and **$P < 0.01$.

Multiple regression analysis. Those factors that correlated significantly with comprehension were considered for two forward, stepwise, multiple regression analyses performed with the aim of investigating if they were collectively predicting comprehension. In the first multiple regression analysis, the number of words, sentences, connectives and co-references were used as predictors and the percentage of correct answers in literal comprehension questions used as a dependent variable. The best regression model was obtained in the first step where only the variable "number of co-references" was entered, accounting for 14% of the variance in the literal comprehension of our students with ID (Adjusted R2 =.14, F (1, 45) =8.42, p <.001; Beta =-.40, p <.01). As the Beta coefficient is negative, this result means that the more co-references there are in the text, the lower the literal comprehension of the students.

In the second multiple regression analysis, the number of words and sentences were used as predictors and the percentage of correct answers in inferential comprehension questions used as a dependent variable. Again, we obtained a significant 1-step regression model where the number of sentences accounted for 12 % of the variance in inferential performance (Adjusted R2 =.12, F (1, 43) =6.74, p <.01; Beta=.37, p<.01). The negative Beta value indicates an inverse relationship between the two factors; i.e. the more propositions there are in the text, the lower the inferential comprehension of students.

Do prior reading skills and IQ levels predict literal and inferential reading comprehension in students with ID?

Finally, in order to examine the relationship between prior reading and IQ levels of students and comprehension performance, several Pearson correlation

analyses were performed between the direct scores in each of the standardized reading tests on the one hand and the percentage of correct answers in literal and inferential comprehension of news on the other. Only direct scores in the standardized reading comprehension test (ECL-1) positively predicted news comprehension at both literal and inferential levels (literal, $r(N=16) = .74$, $p=.001$; inferential, $r(N=16)= .53$, $p= .003$), that is, the higher the ECL-1 score, the higher the literal and inferential comprehension of the news. The rest of correlations were not significant (K-BIT, literal; $r(N=16)= .41$, $p=.12$; inferential $r(N=16)= .39$, $p=.14$; PEABODY, literal, $r(N=16) =.29$, $p=.27$; inferential, $r(N=16) =.05$, $p=.85$; and PROLEC-SE, literal, $r(N=16) =.13$, $p=.62$; inferential, $r(N=16) =.05$, $p=.84$).

Discussion

Reading comprehension level of a corpus of easy-to-read texts by intellectually disabled students

Regarding our first research question, the results indicate that the corpus of easy-to-read texts tested is actually comprehended by our sample of intellectually disabled students, especially at a literal level, with 87% of questions correctly answered. These data are optimistic regarding the use of easy-to-read texts as inputs for students with ID but also suggest that the design method applied to simplify this corpus of texts does not guarantee deep or inferential levels of reading comprehension for such students. In other words, although the average performance in inferential questions was high (76%), 24% of inferential questions were answered incorrectly by our students despite the apparent simplicity of the texts in our corpus, with an average of 94 words. These results agree with those of Karreman et al. (2007) who observed that participants with ID who used a

website adapted following easy-to-read guidelines (Freyhoff et al., 1998) answered more literal and inferential questions correctly than when using a non-adapted version of the website. Although the difference between the number of correct literal and inferential answers was not statistically tested in the study of Karreman et al., they also observed higher scores in literal questions than in inferential questions. Our congruent results support the idea of conducting theoretically driven research that, in our case, made it possible to distinguish between different levels of reading comprehension, based on Kintsch's model.

Surface and deep linguistic measures as predictors of literal and inferential reading comprehension

With regard to the second research question, that is, the relationship between linguistic variables and comprehension performance, we found that only one linguistic variable (number of co-references), predicted literal comprehension, explaining just 14% of its variance. However, the relationship between these two variables was inverse, in contrast to what previous literature seemed to indicate. For instance, Crossley, Greenfield, & McNamara (2008) found that the number of co-references (measured by noun overlaps) was a good predictor of text difficulty (R2=.63); i.e. the higher the number of co-references, the better the comprehension scores in a cloze test. Similar results were found by McNamara et al. (2010): noun co-reference was highly discriminative between low and high-cohesion texts which led to low and high comprehension, respectively. It could be hypothesized that in our texts the number of co-references co-varied with text length, which, in turn, could be explaining the literal comprehension variance. Alternatively, as noted by Meisel (1980), the redundancy introduced in simplified texts by means of noun overlaps would make the sentences´ grammar more

complex and unnatural. As was shown in the reading skills baseline description, our students with ID present a low syntactic proficiency which prevents them from understanding unnatural sentence grammar.

Concerning inferential comprehension, the only linguistic feature that was a good predictor was the number of sentences. The higher the sentence density, the lower the ability of our students with ID to find relationships between sentences. This result could be related to the lower capacity of ID students to store and retrieve information from memory. Actually, Wolman (1991) found that children with ID recalled significantly less statements and episodes from the stories than normally developed children. Therefore, if students with ID have problems in retaining information, increasing the number of information units could hinder the process of making inferences between parts of the text or between the text and prior knowledge. In the end, the impossibility of making inferences, which would concurrently reduce the amount of information units to work with, could cause cognitive overload, leading to students with ID performing a superficial processing of texts.

Alternatively, text length could have affected our students' self-efficacy and reading motivation. Several studies have found that students with ID tend to be reluctant to read long texts (Gómez-Puerta, 2011; Morgan & Moni, 2008) maybe because text length works as a superficial difficulty cue for them. Therefore, their perception of low self-efficacy with respect to comprehension of difficult texts may have led them to actively avoid reading and at best become passive readers (Guthrie & Davis, 2003). If students with ID did not invest enough effort reading long sentences in our study, they probably answered inferential multiple choice questions randomly. Consequently, as suggested by Guthrie and Davis, a primary challenge for the literacy improvement of students with ID would be to re-engage

them. These authors propose that, in addition to providing a wide range of texts focused on content that deals with real-life problems (like the journalist texts used in this research) and the selection and design of texts matched to their reading ability, a pathway to the development of reading motivation is to give direct instruction for important reading comprehension strategies such as questioning, searching, summarizing, using graphic organizers, comprehension monitoring and critical evaluation. Guthrie and Davis (2003)'s rationale for this solution is that:

> As students learn the strategies, they gain the confidence in their own capabilities. They increase their view of themselves as competent in reading. Engagement is fuelled by the self-perception of ability to perform meaningful reading tasks. Consequently, strategy instruction fulfils the motivational need for self-perceived competence (Ryan & Deci, 2000) as well as the cognitive need for possessing skills that are central to text comprehension. (p. 77)

Another interesting result is that related to the effect of connectives: the number of connectives was negatively correlated with literal comprehension, although it did not contribute significantly to our regression model. It means that, probably, its effect is explained as well by a co-variation with other variables such as the length or conceptual density of the text. That could explain the negative effect on comprehension; i.e. the longer the text, the higher the number of connectives and also the number of ideas contained in the text (since connectives are supposedly added to connect ideas). Sanders et al. (2007) did find a facilitative effect of connectives on comprehension, but the conceptual complexity of original and experimental texts was identical in their study; i.e. the only aspect which varied between the two conditions was the presence/absence of connectives. In other words, connectives can facilitate text comprehension but, simultaneously, when they are present in a text it means that such a text contains

ideas that must be integrated. In turn, that would be good news for the education of students with ID because, according to that interpretation, our results would indicate that designers of easy-to-read texts do not entirely renounce using texts with a certain level of conceptual richness. From a motivational point of view, the selection of texts personally relevant and connected to students' needs may compensate for the conceptual richness and complexity of long texts.

It is also worth noting that the effects of word frequency and the Flesch-Szigriszt Index on our measures of comprehension were null; i.e. they did not correlate significantly with either literal or inferential comprehension scores. These results are surprising because these two measures have traditionally been used to select and design texts in educational contexts. According to our data, more complex measures are needed at the sentence and textual cohesion level to ensure that educational text will be understood by ID students.

Individual differences in reading skills of the students with ID

In addition to linguistic variables, the individual differences in reading skills of the students with ID could also be affecting their comprehension performance. Amongst the range of individual measures evaluated (vocabulary knowledge, syntactic skills and comprehension), only the reading comprehension levels of our students positively predicted both literal and inferential comprehension; i.e. those students with better prior levels of reading comprehension took more advantage of the easy-to-read material than students with lower levels. This result suggests that, when adapting texts for this type of students, individual difference in reading comprehension, have to be taken into account.

On the other hand, the lack of predictive power of vocabulary and syntactic skills on reading comprehension could be explained by a floor effect. That is, the levels of vocabulary and syntactic skills were as low in general in our students as not to have an effect on comprehension. Therefore, we cannot discard the idea that given a particular level of performance, those linguistics variables can predict reading comprehension in students with ID. Anyhow, our students' reading profile is similar to the obtained by other researchers (Aitchison, 1994; Moni & Jobling, 2001) which signals the necessity to study it in depth in order to elucidate its origin: acquisition delay, deficit in a particular reading process or lack of reading motivation and engagement. Given the scarcity of specific theories and previous hypotheses about intellectual disability reading comprehension, a more qualitative and inductive methodology such as the case study or the narrative research may help to come up with ideas to study and develop research questions.

Conclusions: methodological shortcomings and future research

In conclusion, we have achieved the two goals of our study; i.e. to validate a corpus of easy-to-read texts for students with ID and to broaden the knowledge of the relationship between texts' linguistics variables, on the one hand, and reading comprehension performance at both literal and deep/inferential level on the other.

However, although our research questions have been answered, further research is needed in order to tackle the several limitations of our study. One of our drawbacks is related to the small sample size (N=16) which, together with the highly selected sample (as opposed to a heterogeneous population of students with ID), is limiting the generalization of our findings. Therefore, before it can be

generalized our study should be replicated with a larger sample size and other levels of intellectual disability. In order to answer the question of qualitative difference versus acquisition delay of particular reading processes suggested by the Perovic (2006)'s findings about syntactic processing in Down syndrome, we would also need a control group formed by typically developed students (without ID) but similar reading comprehension. In fact, the comprehension-age match design has become a popular tool for investigating the causes of reading comprehension failure (Cain, Oakhill, & Bryant, 2000). Let's imagine that we are interested in the role of casual versus temporal connectives (e.g. because vs. when) in texts comprehension. If the typically developed group showed difficulties with temporal but not casual connectives and the reverse pattern was observed in the group of students with ID matched in reading comprehension age, then it could be concluded that connective processing is not delayed in the group of students with ID but qualitatively different with regard to the control group.

A second important methodological limitation that needs to be overcome in future research is the exclusive use of multiple-choice questions as a measure of reading comprehension which could be preventing students from making and reporting their own inferences. Actually, Karreman et al. (2007) observed that open questions were more sensitive to comprehension differences between students with ID than multiple-choice questions. However, the risk of using open questions is that memory and comprehension skill can be confounded. In addition to the memory issue, as it was highlighted in the material section, multiple-choice questions were considered more suitable for this particular study than open questions due to several reasons: the prevention of the potential masking effects of writing competence or verbalization difficulties of our students (Fletcher & Clayton, 1994); the facilitation of readers' answer coding (of a total of 178

questions); and the activation of specific types of reading processes (literal vs. Inferential) in our readers. If these constraints are not present in the assessment situation (e.g. writing and verbalization difficulties have been discarded), then we encourage educators to use open questions instead of close ones in order to detect unexpected inferences and elaborations made by students.

A third methodological issue is related to the correlational design of our study which has allowed us to describe the reading comprehension behaviour of a group of students with ID and to identify several text linguistic variables related to it but it does not allow us to answer causal questions. Therefore, in order to study in depth the origin of our findings such us co-reference or text length effects on literal and inferential comprehension respectively, process measures like eye movements could be added to output measures (e.g. comprehension questions). Eye movement measures could help us to detect if readers are actually reading long sentences or neglecting them and paying especial attention (longer fixation times and revisits) to particular elements of long texts such as connectives or co-references. This type of process measure, combined with an inductive methodology like a case study, could provide us with an account of causal hypotheses to be tested by means of comprehension-age match designs.

Finally, with the aim of augmenting reading engagement, future research could focus on the use of electronic texts since the Internet and multimedia environments have been found to be highly motivating for youth students with ID (Gómez-Puerta, 2011). However, we wanted to note that the use of electronic texts is not an un-risky engagement solution. Electronic reading not only involves reading ability but other kind of skills such as navigation and computer operating skills (e.g. operating a mouse, typewriting, etc., for a revision see Van Deursen,

van Dijk, & Peters, 2011) which paradoxically might make more difficult the reading comprehension process for our target students.

Despite the limitations of our methodology, we think that our findings throw light on the validity of superficial and deep measures of text difficulty for students with ID. From an applied point of view, we expect our insights to assist professionals in the field of education (teachers, text designers, etc.) in the task of more optimally matching texts and ID students´ reading levels.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times. *Psychological Science*, *17*(9), 814–823. http://doi.org/10.1111/j.1467-9280.2006.01787.x

Aitchison, J. (1994). *Words in the Mind: An Introduction to the Mental Lexicon* (3rd ed.). Oxford: Blackwell.

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

American Psychiatric Association. (2000). *DSM-IV*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Association.

Baayen R. H., Piepenbrock R. & Gulikers L. (1995) *The CELEX Lexical Database*. University of Pennsylvania, Linguistic Data Consortium, Philadelphia.

Bos C. & Tierney R. (1984) Inferential reading abilities of mildly mentally retarded and nonretarded students. *American Journal of Mental Deficiency* 89, 75–82.

Cain, K., Oakhill, J. V., & Bryant, P. E. (2000). Investigating the causes of reading comprehension failure: the comprehension-age match design. *Reading and Writing: An Interdisciplinary Journal*, *12*, 31–40. http://doi.org/10.1023/A:1008058319399

Conners, F. A. (2003). Reading skills and cognitive abilities of individuals with mental retardation. In Abbeduto (Ed.), *International review of research in mental retardation* (27th ed., pp. 191–229). New York: Academic Press.

Conners, F. A., Atwell, J. A., Rosenquist, C. J., & Sligh, A. C. (2001). Abilities underlying decoding differences in children with intellectual disability. *Journal of Intellectual Disability Research : JIDR*, *45*(Pt 4), 292–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11489050

Crossley, S. a., Dufty, D. F., McCarthy, P. M., & McNamara, D. S. (2007). Toward a new readability: A mixed model approach. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 197–202. Retrieved from http://www2.gsu.edu/~wwwesl/Files/ALSL/Cros_Toward_a_new_readability_A_mixed_model_approach.pdf

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, *42*, 475–493. http://doi.org/10.1002/j.1545-7249.2008.tb00142.x

De la Cruz, M. (1999). *Evaluación de la Comprensión Lectora (ECL-1)*. Madrid: TEA.

Doctorow, M., Wittrock, M. C., & Marks, C. (1974). Generative processes in reading comprehension. *Journal of Educational Psychology*. http://doi.org/10.1037/0022-0663.70.2.109

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III: Peabody test de vocabulario en imágenes*. Madrid: TEA.

Ferrer, A., Brancal, M., & Asensi, Mªc. (1999). La evaluación de la producción léxica en las personas con retraso mental: alternativas a la medición psicométrica. In *III Congreso Internacional de Psicología Evolutiva y de la Educación*. Santiago de Compostela.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, *32*(3), 221–233.

Fletcher, J., & Clayton, I. (1994). Measuring listening comprehension in adolescents with intellectual disability. *Australia New Zealand Journal of Developmental Disabilities*, *19*(1), 53–59.

Fowler A. E., Doherty B. J. & Boynton L. (1995) The basis of reading skill in young adults with Down syndrome. In: *Down Syndrome: Living and Learning in the Community* (eds L. Nadel & D. Rosenthal), pp. 182–196. Wiley-Liss, New York.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 361–379. http://doi.org/10.1037/0096-1523.2.3.361

Freebody, P., & Anderson, R. C. (1983). Effects of Vocabulary Difficulty, Text Cohesion, and Schema Availability on Reading Comprehension. *Reading Research Quarterly*, *18*(3), 277–294. http://doi.org/10.2307/747389

Freyhoff, G., Hess, G., Kerr, L., Menzel, E., Tronbacke, B., & Der, VanVeken, K. (1998). *Make it simple: European guidelines for the production of easy-to-read information for people with learning disability for authors, editors, information providers, translators and other interested persons*. ILSMH European Association.

Gómez-Puerta, M. (2011) Identificación de dificultades de comprensión lectora en el uso de Internet en jóvenes con discapacidad intelectual para el desarrollo de un periódico digital. Poster presented at the XV Congreso Nacional y I

Internacional de Modelos de Investigación Educativa, Madrid, 2011, September.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(2), 193–202. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15354684

Guthrie, J. T., & Davis, M. H. (2003). Motivating Struggling Readers in Middle School Through an Engagement Model of Classroom Practice. *Reading & Writing Quarterly*, *19*(1), 59–85. http://doi.org/10.1080/10573560308203

Jones, F. W., Long, K., & Finlay, W. M. L. (2007). Symbols can improve the reading comprehension of adults with learning disabilities. *Journal of Intellectual Disability Research : JIDR*, *51*(Pt 7), 545–50. http://doi.org/10.1111/j.1365-2788.2006.00926.x

Karreman, J., Van Der Geest, T., & Buursink, E. (2007). Accessible website content guidelines for users with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *20*, 510–518. http://doi.org/10.1111/j.1468-3148.2006.00353.x

Kaufman, A. S., & Kaufman, N. L. (1997). *Test breve de inteligencia de Kaufman (K-BIT)*. Madrid: TEA.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. http://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. http://doi.org/10.1037/0033-295x.85.5.363

Lee W. Y. (1995) Authenticity revisited: text authenticity and learner authenticity. *ELT Journal* 49, 323–328.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes*. http://doi.org/10.1080/01638530902959943

Meisel, J. (1980). Linguistic simplification. In S. Felix (Ed.), *Second Language: Trends and Issues* (pp. 13–140). Tübingen: Gunter Narr.

Moni, K. B., & Jobling, A. (2001). Reading-related Literacy Learning of Young Adults with Down Syndrome: Findings from a three year teaching and research program. *International Journal of Disability, Development and Education*, *48*(4), 377–394. http://doi.org/10.1080/10349120120094275

Monsalve A. & Cuetos F. (2001) Asimetría hemisférica en el reconocimiento de palabras: efectos de frecuencia e imaginabilidad. *Psicothema* 13, 24–28.

Morgan, M. F., & Moni, K. B. (2008). Meeting the challenge of limited literacy resources for adolescents and adults with intellectual disabilities. *British Journal of Special Education*, *35*(2), 92–101. http://doi.org/10.1111/j.1467-8578.2008.00378.x

Nomura, M., Nielsen, G. S., & Tronbacke, B. (2010). *Guidelines for easy-to-read materials: Revision*. The Hague: IFLA Headquarters.

Perovic, A. (2006). Syntactic deficit in Down syndrome: More evidence for the modular organisation of language. *Lingua*, *116*(10), 1616–1630. http://doi.org/10.1016/j.lingua.2005.05.011

Poncelas, A., & Murphy, G. (2007). Accessible Information for People with Intellectual Disabilities: Do Symbols Really Help? *Journal of Applied Research in Intellectual Disabilities*, *20*(5), 466–474. http://doi.org/10.1111/j.1468-3148.2006.00334.x

Ramos, J. L., & Cuetos, F. (1999). *PROLEC-SE: evaluación de los procesos lectores en alumnos del tercer ciclo de educación primaria y educación secundaria obligatoria*. Madrid: TEA.

Rondal, J. A. (2001). El lenguaje en el retraso mental: diferencias individuales y sindrómicas y variación neurogenética. *Revista de Logopedia, Foniatría Y Audiología*, *21*(2), 46–63. http://doi.org/10.1016/S0214-4603(01)76185-2

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. http://doi.org/10.1006/ceps.1999.1020

Ryder R. & Hughes M. (1985) The effect on text comprehension of word frequency. *The Journal of Educational Research* 78, 286–291.

Sanders T., Land J. & Mulder G. (2007) Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal* 15, 219–235.

Szigriszt-Pazos, F. (1992). *Sistemas predictivos de legilibilidad del mensaje escrito: fórmula de perspicuidad*. Universidad Complutense de Madrid. Servicio de Publicaciones.

Tronbacke, B. (1997). *Guidelines for Easy-to-read Materials*. (I. Headquarters, Ed.).

Van Deursen, A. J. A. M., van Dijk, J. A. G. M., & Peters, O. (2011). Rethinking Internet skills: The contribution of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. *Poetics*, *39*(2), 125–144. http://doi.org/10.1016/j.poetic.2011.02.001

Verhoeven, L., & Vermeer, A. (2006). Literacy achievement of children with intellectual disabilities and differing linguistic backgrounds. *Journal of Intellectual Disability Research : JIDR*, *50*(Pt 10), 725–38. http://doi.org/10.1111/j.1365-2788.2006.00838.x

Wolman, C. (1991). Sensitivity to Causal Cohesion in Stories by Children with Mild Mental Retardation, Children with Learning Disabilities, and Children without Disabilities. *The Journal of Special Education*, *25*(2), 135–154. http://doi.org/10.1177/002246699102500202

Appendix A

Example of one of the journalist texts used in this study and two of the literal and inferential questions which accompanied it (originals in Spanish).

Tickets for the match Real Madrid–Alcorcón are sold out

The football match between the Real Madrid and the Alcorcón which will take place next Tuesday has piqued the interest of everybody and all tickets were sold out today. The stadium of the Real Madrid will be full. The result of the last match in the Alcorcón stadium, where Real Madrid lost 4-0, is the reason why this match is so interesting.

Comprehension questions:

1. When will the match Real Madrid–Alcorcón take place? **Literal**
   a. Next month.
   b. Next Tuesday. **Correct**
   c. Today.

2. Why are the tickets for the match Real Madrid–Alcorcón sold out? **Inferential**
   a. Because the Alcorcón won Real Madrid 4-0 in the last match. **Correct**
   b. Because the pitch of the Real Madrid will be full.
   c. Because the match will be on Tuesday.

4 Study 2

Towards text simplification for poor readers with intellectual disability: When do connectives enhance text cohesion?

Fajardo, I., Tavares, G., Ávila, V., & Ferrer, A. (2013). Towards text simplification for poor readers with intellectual disability: when do connectives enhance text cohesion? *Research in Developmental Disabilities: RIDD*, 34(4), 1267–79. http://doi.org/10.1016/j.ridd.2013.01.006

Towards text simplification for poor readers with intellectual disability: When do connectives enhance text cohesion?

Inmaculada Fajardo*, Gema Tavares, Vicenta Ávila, Antonio Ferrer

*ERI Lectura, University of Valencia

Abstract

Cohesive elements of texts such as connectives (e.g., but, in contrast) are expected to facilitate inferential comprehension in poor readers. Two experiments tested this prediction in poor readers with intellectual disability (ID) by: (a) comparing literal and inferential text comprehension of texts with and without connectives and/or high frequency content words (Experiment 1) and (b) exploring the effects of type and familiarity of connectives on two-clause text comprehension by means of a cloze task (Experiment 2). Neither the addition of high frequency content words nor connectives in general produced inferential comprehension improvements. However, although readers with ID were less likely to select the target connective in the cloze task than chronologically age-matched readers (mean age = 21 years) in general, their performance was affected by the type of connective and its familiarity. Familiarity had a facilitative effect for additive and contrastive connectives, but interfered in the case of temporal and causal connectives. The average performance of a reading level-matched control group (typically developing children) was similar to the group of readers with ID although the pattern of interaction between familiarity and type of connectives varied between groups. The implications of these findings for the adaptation of texts in special education contexts are discussed.

Keywords: Word frequency, connectives, literal and inferential text comprehension, intellectual disability.

Towards text simplification for poor readers with intellectual disability:

When do connectives enhance text cohesion?

Literature review

It is well established that individuals with intellectual disability (ID) present reading difficulties at both decoding and comprehension levels (see Conners, 2003 for a review). Despite intervention efforts in special education settings to improve literacy in students with ID, several studies have recently indicated that the average level of reading and comprehension abilities of students with ID (aged 16–22) is equivalent to that observed in primary students in the age range of 6–10 years (Fajardo, Ávila, Tavares & Ferrer, 2010; Moni & Jobling, 2001; Morgan & Moni, 2008). This reading literacy problem is dramatically constraining their academic success, access to information (e.g., news), job opportunities (e.g., access to job offers) and even their entertainment options (use of digital social networks, forums, etc.).

In academic settings, the instructional intervention approach, aimed at directly improving reading abilities (e.g., Alberto, Waugh, & Fredrick, 2010; Allor, Mathes, Champlin, & Cheatham, 2009; Cohen et al., 2006; Gersten, Fuchs, Williams, & Baker, 2001; Joseph & Eveleigh, 2011; Van der Bijl, Alant, & Lloyd, 2006), has been used alone or in combination with the text simplification approach. The second approach is characterised by the modification of the text in order to make it more legible for a target reader group. Both approaches seem to be essential for making reading activities not only profitable but enjoyable for readers with special needs (Mastropieri & Scruggs, 1992).

Therefore, the deliberate simplification of texts is a common practice for teachers dealing with adolescents and adults with intellectual disabilities (Morgan & Moni, 2008). The interest in this type of approach is represented by official educational programs like the one framing the study presented here. The program, called ‘‘Lectura Fácil’’ (funded by the Spanish Ministry of Industry in 2009), ‘‘Easy-Reading’’ in English, is aimed at promoting newspaper reading in students with ID. As part of the program, news articles within the range of interests of students with ID are selected on a daily basis, simplified and uploaded on a website especially designed for this aim, which has been operating since 2009 (http:// www.noticiasfacil.es). The general aim of the first experiment of the study reported here was to measure the utility of the program by comparing different ways of simplifying the news articles to improve their readability (how easy a text can be read and understood).

Text simplification is not a straightforward issue. In order to simplify, adapt, create or simply select texts that match students’ reading level, several factors can be considered, such as interest and relevance for the learner, use of illustrations, orthotypography issues or linguistic features of the texts (Morgan & Moni, 2008; Tronbacke, 1997). The purpose of the present study was to bring the latter ones, that is, text linguistic features, into focus by identifying, which and how they might influence text comprehension for students with ID.

From a linguistic point of view, Crossley, Louwerse, McCarthy, and McNamara (2007) suggest two approaches to evaluate text readability: (a) the use of shallow-based readability measures and (b) the use of textual cohesion measures (also called deep readability measures). Following the framework of Kintsch’s construction-integration model (Kintsch, 1988), each approach would be related to different levels of reading comprehension. On the one hand, shallow-

based readability measures, for example, word length (number of characters or syllables), sentence length (number of words) and word frequency (number of appearances of a word in a corpus of texts) would mainly affect the literal comprehension of texts, that is, comprehension of the strict meaning of single propositions. On the other hand, textual cohesion features (presence or density of connectives- linking devices such as *but*, *and* or *for that reason* and co-references – anaphors such as pronouns or repeated names in a text) would affect the inferential level, that is, the integration between text segments or between text and prior knowledge.

In the following sections, the literature about the influence of word frequency and presence of connectives (a shallow- based and a cohesion linguistic feature respectively) on reading comprehension performance of students with and without ID is briefly reviewed. Afterwards, the research goals and hypotheses of Experiment 1 are stated.

The effect of word frequency on reading comprehension.

Traditional shallow-based readability measures assume that the shorter the word and sentence length (measured by means of standard formulas such as the Flesch Index by Flesch, 1948) and the higher the word frequency [e.g., using word frequency databases such as CELEX by Baayen, Piepenbrock, and Gulikers (1995) for English or Alameda and Cuetos (Alameda & Cuetos, 1995) for Spanish], the lower the text difficulty and the higher the readers' comprehension. However, empirical evidence does not provide clear support for this assumption, especially with regard to word frequency. While some studies have found a facilitative effect of word frequency on reading comprehension (Doctorow, Wittrock, & Marks, 1974;

Ozuru, Rowe, O'Reilly, & McNamara, 2008), others studies have shown no effect for readers with a regular level of reading skills (Freebody & Anderson, 1983; Ryder & Hughes, 1985).

When it comes to students with ID, who tend to present a small receptive vocabulary (amount of words a person recognises and understands when read or heard), which is strongly correlated with their reading comprehension performance (Nash & Heath, 2011), we would expect that texts with high word frequency would certainly enhance literal comprehension. For instance, in the sentence ''Anne and Tom spent an agreeable afternoon'', the substitution of the term ''agreeable'' with the term ''lovely'' would theoretically make the sentence easier to understand for a person with an equivalent vocabulary age of 6 years since ''lovely'' is a more frequent word than ''agreeable''.

To our knowledge, there are only two previous studies that have addressed this issue in students with ID showing, as in the case of regular readers, contradictory findings. On the one hand, Karreman, der Geest, and Buursink (2007) adapted several digital texts (texts included on a website) by modifying different linguistic elements (e.g., word and sentence length, frequency and abstractness of words) and asked groups of students with ID to answer comprehension questions after reading both versions (between-subject manipulation). The results showed that both literal and inferential comprehension were higher in the adapted version than in the non-adapted version of the same digital texts. However, as word frequency was manipulated at the same time as other linguistic features (word length, sentence length, etc.) in the adapted texts, it is unclear whether word frequency actually contributed to the comprehension and, if so, to what extent.

On the other hand, in a previous study conducted by the authors of this manuscript within the program ‘‘Lectura Fácil’’ (Fajardo et al., 2010) students with mild ID were asked to read a set of 48 journalistic texts (during a period of 16 weeks) and answer literal and inferential questions about them afterwards. In this case, participants did not read adapted and non- adapted versions of the same texts like in Karreman et al.’s (2007) study. It was assumed instead that the 48 texts varied in the target linguistic variables (e.g., word frequency, word and sentence length, etc.), which were measured and correlated with students’ comprehension scores. Contrary to the word frequency facilitation assumption, there was no significant positive correlation between word frequency and comprehension (at either the literal or inferential level). Using the dictionary of frequencies of the Spanish linguistic units (Alameda & Cuetos, 1995), a database of two million words, some studies have considered a minimum frequency of occurrence of 20 per two-million as the high frequency boundary for regular readers (e.g., Alvarez, Carreiras, & Taft, 2001). Therefore, in Fajardo et al.’s study (2010), presumably there was a ceiling effect of this variable since the average word frequency of the 48 texts of the study was very high (average frequency of 21,361 [*SD* = 4427] occurrences per two million words according to Alameda and Cuetos’ database). That means that given a specific level, word frequency would no longer have an effect on reading comprehension. However, which is the minimum critical level of word frequency for students with ID? As we will explain later, this question is related to the first goal of Experiment 1, that is, to test whether text word frequency matched to the receptive vocabulary level of this kind of students would improve their text reading comprehension at the literal level.

Finally, with the aim of augmenting the probability that text vocabulary is known by readers, direct vocabulary training is presented as an alternative to text

word frequency increment. If training improved reading comprehension, it could be taken as indirect evidence of the facilitative effect of word frequency on text comprehension. Therefore, to conclude this section, we should mention the study of Pany, Jenkins, and Schreck (1982) who observed that direct vocabulary training in students with learning disabilities transferred to comprehension of single sentences. However, no effects attributable to vocabulary instruction on passage comprehension (measured by global and inferential measures such as a cloze test – a short text with blanks where some of the words should be and readers had to fill by integrating parts of the text – and a story-retell test – whereby readers were asked to tell everything they could remember about the story and the number and complexity of inferences made by them were analysed) were observed. This finding suggests that the understanding of the literal meaning of sentences enhanced by vocabulary knowledge does not guarantee higher levels of text reading comprehension for which inferential reasoning is required (Deane, Sheehan, Sabatini, Futagi, & Kostin, 2006). In the next section, the literature about the use of connectives to increase inferential comprehension is discussed.

The effect of connectives on reading comprehension.

As we said above, connectives are linking or cohesive features that would affect the inferential level of reading comprehension, that is, the integration between text segments or between text and prior knowledge. A connective (e.g., but, in addition, because) is characterised as a lexical unit that makes sentence relations explicit. For instance, the causal connective *because* marks the cause–effect relationship between the current clause and, usually, the previous one in the sentence. The reactivation hypothesis of Millis and Just (1994) proposes that the presence of connectives increases the activation level of the content of the

preceding clause, thus facilitating inter-clause integration. This mechanism could explain the findings of Sanders, Land, and Mulder (2007) who observed that individuals with poor comprehension levels (non-ID) obtained higher comprehension scores after reading texts with connectives than texts without them.

In typically developing children (aged 8–10 years), connectives seem to aid text processing as well. For instance, Cain and Nash (2011) observed that, although knowledge and comprehension of connectives is still developing in young readers, they are able to read texts more quickly when two-clause sentences are linked by an appropriate connective compared with texts with an inappropriate or not present connective.

Then, could we expect a facilitative effect of connectives in poor readers with ID? There are few previous experiments about the effect of connectives on ID readers' text comprehension. The study of Verhoeven and Vermeer (2006) measures the ability of children with ID to infer connectives and other types of linking devices (e.g., pronouns). In this study, children with ID were asked to complete text fragments in which those linguistic elements had been omitted (the typical cloze task). For example, in the sentence Daniel put on his coat before going out. He did it _____ it was raining, the missing connective is *because*. The results showed that the scores of children with ID in this task were significantly lower than those of typically developed children.

Reading ability-related differences in the use of connectives by typically developed children have also been observed. Cain, Patson, and Andrews (2005) found that children with poor reading comprehension levels (aged 8–9 years) were less likely to select the correct target connective in a cloze task than good comprehenders.

One possible explanation for this finding is that poor readers find it difficult to understand (to infer) the type of relationship established between sentences (e.g., causality or temporality) and consequently they are unable to correctly guess the connective that should accompany them. This seems to suggest that the presence of connectives in the text could help the comprehension process by making explicit the relation between clauses (e.g., by increasing the activation level of the preceding clause).

However, negative effects of connectives have also been reported in the literature. For instance, Noordman, Vonk, and Kempff (1992) observed that the presence of connectives tends to speed up reading time, but interfere with the inferential process by preventing readers from activating prior knowledge. In other words, connectives would make explicit the relationship between sentences (e.g., causal, adversative, etc.), increasing the passivity of readers during the reading process. Another example of the negative impact of connectives on reading comprehension comes from our own research. In the already mentioned Fajardo et al. (2010) study, deep readability variables such as number of connectives and co-references in the sample of 48 texts read by students with ID were also measured. We observed that the number of connectives was negatively correlated with literal comprehension, in other words, the higher the number of connectives the lower the comprehension of single sentences. Nevertheless, in this study the number of connectives co-varied with other variables such as the number of sentences of the text. The negative effect on comprehension could be explained as follows: the longer the text, the number of ideas contained in the text and also the higher the number of connectives needed to link these ideas.

In other words, connectives could facilitate inter-clause integration but, simultaneously, their presence in a text would mean that it contains a higher

number of ideas that must be recalled, processed and locally (literally) comprehended.

Therefore, it seems that the manipulation of connectives, as it might occur with other types of linguistic features, rarely has an isolated effect on comprehension. In the next section the effect of simultaneously combining the two linguistic features discussed up to now, word frequency and connectivity is addressed.

Combined effect of word frequency and connectivity on comprehension.

A priori, independent effects of each variable could be hypothesised since each one would affect a different level of reading comprehension, as Crossley et al.'s (2007) classification of text readability measures suggests. A text with high frequency words and connectives would facilitate the literal understanding of sentences and the connection between them (inferential comprehension) respectively. Paradoxically, the traditional text simplification approach has made the manipulation of shallow and deep features interdependent, for instance, by eliding connectives (deep measure) in order to reduce sentence length (shallow measure). Consequently, some researchers (e.g., Goodman & Freeman, 1993; Long & Ross, 1993) have argued that simplified texts lack the cohesiveness of authentic texts (because of the absence of connectives) and as a result are more difficult to comprehend and decipher than the latter.

As the empirical literature is not conclusive, we designed Experiment 1 to examine the isolated and combined effects of word frequency and connectivity on literal and inferential comprehension in students with ID. In particular, two linguistic features, word frequency and connectives, classified as shallow and deep

(cohesion) readability features respectively, were used both separately and combined to simplify a set of texts. The effect of these two conditions on literal and inferential comprehension was compared to non-adapted texts in students with mild ID. Next, the particular hypotheses and the methodology used to explore this issue are described.

Experiment 1

Hypotheses

H1. Text simplification based on increasing word frequency (frequency adaptation) by means of the substitution of low frequency words with high frequency words (according to the receptive vocabulary level of readers) will improve literal comprehension for students with ID compared to original texts.

H2. Text simplification based on increasing text connectivity (connectivity adaptation) by means of adding connectives between sentences where necessary will improve inferential comprehension for students with ID compared to original texts.

H3. Text simplification based on frequency plus connectivity (combined adaptation) will improve both literal and inferential comprehension in students with ID compared to original texts, that is, everything else controlled; an independent effect of each variable is expected when both frequency and connectivity adaptations are implemented simultaneously.

Method

Participants. Nineteen students (nine males, ten females with an average age of 19, *SD* = 2) recruited from Camí Obert, a vocational training centre for

students with ID belonging to the Asociación Pro-Discapacitados Psíquicos de Alicante (Spain) participated in this study. Students signed an informed consent form prior to their participation in the study.

Measures of students' IQ and baseline reading abilities. The selection criteria to participate in this experiment were two-fold: (a) borderline to mild intellectual disabilities and (b) reading comprehension level equivalent to the third grade of elementary school (ages eight to nine). Third grade students in Spanish curricula are asked to work on text comprehension (reading longer texts than in previous grades and distinguishing between a variety of genres) so they are supposed to have basic decoding skills. Participants who fulfilled this criterion were supposed to be able to manage the reading task used in this experiment.

In what follows, characteristics and results of the IQ and reading comprehension tests are described. Two additional measures obtained to set a baseline of abilities related to reading are also reported.

Intelligence. The Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1997) was used to measure the verbal and non-verbal intelligence of the students. It is a brief, individually administered measure of the intelligence of a wide range of individuals spanning the ages of 4–90 years. It is composed of two subtests: Vocabulary (Expressive Vocabulary and Definitions) and Matrices. It takes 15–30 min to administer. The IQ composite standard score (a combination of vocabulary and matrices scores) of the group was 71 (*SD* = 15, 3rd percentile), which corresponds to mild intellectual disability according to the classification of the DSM-IV-TR Manual (American Psychiatric Association, 2000).

Reading comprehension. This was measured through the Reading Comprehension Test 1 (ECL1, De la Cruz, 1999), which consists of three short texts (approx. average length of 100 words) accompanied by literal and inferential questions. The average direct score was 9.8 (*SD* = 2.7), which corresponds to the 65th percentile for the 3rd grade level.

Receptive vocabulary. Receptive vocabulary was measured through the Peabody Picture Vocabulary Test-III (PPVT; Dunn, Dunn, & Arribas, 2006). Students were asked to match a spoken word with one of four pictures. The average direct score was 116.7 (*SD* = 25, 1st percentile), which corresponds to an equivalent age of 10 years and 1 month (*SD* = 3 years and 2 months). Syntactic knowledge. The syntactic knowledge was measured through the syntactic subtest of the PROLEC-SE (Ramos & Cuetos, 1999). It consists of sets of pictures that represent scenes (for instance, a policeman arresting a burglar). Participants were asked to select one written sentence out of four that correctly describes the scene. The syntactic complexity of the sentences increases as trials are displayed, so that the syntactic competence of participants can be established. The average direct score in this subtest was 14 (*SD* = 4.5, 5th percentile).

In summary, the 19 students fulfilled the first inclusion criterion, that is, they presented mild intellectual disability. With regard to reading abilities, the group was below its chronological age in all tests: they fall below the 5th percentile in the PPVT and PROLEC-SE tests, which means that their vocabulary and syntactic knowledge was very poor. However, they had acquired, at least, a minimum level of reading comprehension (equivalent to 3rd grade level) as shown in the reading comprehension test, so they also fulfilled our second criterion to participate in this experiment.

Texts and comprehension questions. Participants were asked to read four versions of short Spanish journalistic texts and answer literal and inferential comprehension questions after each one.

Journalistic texts. The four journalistic texts were selected from real Spanish newspapers (see an example text in Spanish in Appendix A). With the aim of improving generalizability, the topics ranged from sports to economy. To ensure that texts were of similar difficulty, a word limit of 250 words was established (the average number of words was 214, range 205–221). The Fernández-Huerta Index (Fernández-Huerta, 1959), a Spanish version of the classical Flesch index for measuring text difficulty, was used as a measure of external validity. The Fernández-Huerta Index is expressed as follows: 206,84 - (60 x (S/ P)) - (1.02 x (P/F)), where S is the number of syllables, P is the number of words in the text and F is the number of sentences (defined as the content between two punctuation marks). The Fernández-Huerta scale runs from 0 (practically unreadable) to 100 (easy to read). The average Fernández-Huerta Index of our four texts was 62 (range = 55–66).

Three adapted versions of each text were created in addition to the original non-adapted versions: frequency adaptation, connectivity adaptation and combined adaptation. That is, there were four versions of each text. In what follows the four versions are described (see an example text in Appendix A):

(1) Non-adapted version: the texts were used in their original form. The selection criteria were: (a) length limit of 250 words and (b) no presence of between-sentence connectives (although some of the selected texts did contain within sentence connectives such as *and* or *or*).

(2) Frequency adaptation: low frequency words were substituted with high frequency words in the non-adapted texts. Words were considered of low frequency when their occurrence per two million words was below 53 (according to Alameda and Cuetos' 1995 database). Words were considered of high frequency when their occurrence per two million words was above 53. This frequency limit was the average frequency of the words corresponding to the age-equivalent vocabulary level of our participants as measured with the PPVT-III. For instance, the low frequency word *calificaciones* (the Spanish term for *scores* with a frequency of 2) was substituted with *notas* (the high frequency Spanish synonym of *scores* with a frequency of 78).

(3) Connectivity adaptation: connectives between sentences (e.g., however, for that reason) were added to the non-adapted texts. For instance, the connective *however* was added between these two sentences: [. . .] *one of the companies has announced a drastic measure. However, the rest of the companies opted for more progressive measures* [. . .].

(4) Combined adaptation: both frequency and connectivity adaptation were simultaneously applied to the non-adapted texts.

Comprehension questions. After reading each text, participants were asked to answer literal and inferential multiple- choice questions with three answer alternatives (see examples in Appendix A). For literal questions, the answer was either explicitly stated or restructured within a single sentence. The number of literal questions per text was three, except for one of the texts for which four questions were designed. Inferential questions required the integration of information between- sentences. There were four of this type for each text.

Design. The design was a repeated measures Latin-square design to counterbalance experimental conditions (adaptation version) across texts (sport, economy, technology and society). We constructed four lists of texts, with one version of each text appearing in each list. The total number of texts per list was four. Participants were randomly assigned to one of the four lists. Thus, each participant read each text and each experimental condition (non-adapted, frequency adapted, connectivity adapted and combined), but never saw more than one version of the same item.

Procedure. Each participant read four texts in total and answered from seven to eight comprehension questions after reading each one. In order for participants to get familiar with the procedure, they performed a practice trial with a different text before reading the four experimental texts. Afterwards, one of the four lists of texts was randomly assigned to each participant. Within each list, texts were presented in a randomised order on a computer screen and participants were told to click the button questions as soon as they finished reading. Each question and its three answer choices were presented sequentially on separate screens.

Participants were allowed to move between a question and the text until they gave the answer. Once the answer was given, the next question was presented and participants could not go back to the previous one. When the questions on a text had been answered, the participants were asked to read the next text. There was no time limit for completing the task. The software controlling the administration of the experiment and registration of the participants' answers was written in Visual Basic by the researchers.

Results

The descriptive results for the percentage of correct answers for literal and inferential questions for each condition are shown in Table 1.

Table 1.

Experiment 1. Reading comprehension scores for each condition of text adaptation.

	Percentage of correct answers M (*SD*, *Mdn*)	
Type of adaptation	Literal	Inferential
Word frequency	.77 (.30, 1)	.51 (.29, .50)
Connectivity	.79 (.23, .67)	.47 (.23, .50)
Combined	.72 (.30, .67)	.61 (.19, .60)
Non-adapted	.77 (.25, .67)	.47 (.28, .50)

As the sample size was small and the distributions of dependent variables were not normal, data were analysed using the Wilcoxon matched pairs non-parametric test. Firstly, the non-adapted condition was compared with each of the adaptation conditions per level of comprehension (medians for each group are shown in Table 1). Literal comprehension did not differ significantly between any of the adapted and the non-adapted versions of the texts. Regarding inferential comprehension, there was no effect of type of adaptation, that is, none of the adapted versions of the texts obtained significantly higher scores than the non-adapted condition.

Additionally, Wilcoxon matched pairs tests were performed in order to compare literal and inferential comprehension for each adaptation level. Literal comprehension was significantly higher than inferential comprehension in all

experimental conditions (word frequency, $Z = 2.5$, $p = .01$; connectivity, $Z = 3.5$, $p = .001$; combined, $Z = 3.8$, $p = .001$; non-adapted, $Z = 3.03$, $p = .002$).

Discussion

This study aimed to explore the isolated and combined effects of two linguistic features, word frequency and connectivity, considered as shallow and cohesion measures respectively, in literal and inferential reading comprehension of students with ID.

The absence of isolated effects of word frequency agrees with previously observed findings in readers with a regular level of reading skills (Freebody & Anderson, 1983; Ryder & Hughes, 1985) and readers with ID (Fajardo et al., 2010). However, this result contrasts with the facilitative effects found by other authors and hypothesised by us (H1). In the area of vocabulary training, for instance, Pany et al. (1982) observed that the modality of direct training in students with learning disability transferred to comprehension of single sentences, called literal comprehension in the present work. Certainly, vocabulary training and the substitution of low frequency words with high frequency words are quite different approaches to improve reading comprehension. The use of high frequency words does not guarantee that students know the word meaning, while direct vocabulary training does. As some authors have noted before (see discussion in Crossley et al., 2007), the fact that high frequency words in languages such as English or Spanish tend to be polysemous could have meant that adapted versions of the texts were more ambiguous as a whole.

Regarding connectivity, again no isolated effect was found either at the literal level or the inferential level. This result disagrees with previous findings (Cain

& Nash, 2011; Sanders et al., 2007) and the reactivation hypothesis of Millis and Just (1994), which proposes that a connective increases the activation level of the content of the preceding clause, thus facilitating between-sentence integration (H2). Obviously, if independent effects of connectivity and frequency did not appear, the additive effect of both variables predicted by H3 is not supported either.

There are two possible explanations for the lack of facilitative effect of connectivity: (a) the presence of connectives induces a passive processing of texts, preventing the activation of prior knowledge (Noordman et al., 1992) and (b) connective facilitation could be restricted to certain types of connectives. Although both explanations are not incompatible, it seems that there is a larger corpus of research supporting the second one (e.g., Crosson, Lesaux, & Martiniello, 2008; Geva, 2006; McNamara, Louwerse, McCarthy, & Graesser, 2010; Shapiro & Hudson, 1991). For instance, McNamara et al. compared high- versus low-cohesion versions of texts extracted from published discourse psychology studies. In all cases, high cohesion versions improved reading comprehension (measured differently in each study). McNamara et al. calculated the connective incidence by type (additive, causal, temporal and clarification connectives) finding that high-cohesion texts included more causal connectives than low-cohesion texts. However, regarding the other types of connectives, there were no differences between the texts, which suggests that only enhanced causality contributes to text coherence and consequently to reading comprehension improvements. This conclusion must be taken cautiously though, since the authors selected a corpus of texts that were specifically designed to provide additional causal connections between the ideas, so the study presented a bias towards this type of connectivity.

Leaving aside this methodological issue, as we said, there is further evidence for the effect of connective type on comprehension (oral or written)

coming from the field of language development and second language acquisition. Regarding language development, the literature signals that the developmental acquisition order for connectives is as follows: additive, contrastive, temporal, causal, and adversative (e.g., Bloom, Lahey, Hood, Lifter, & Fiess, 1980; Geva, 2006; Shapiro & Hudson, 1991).This order seems related to the complexity of the semantic relation between sentences or segments linked by the connective. Such a complexity would derive from two dimensions (Spooren & Sanders, 2008): basic operation (segments can either be connected strongly – in causal relations – or weakly – in additive relations) and polarity (in a positive relation, the segments are linked directly, while in a negative relation the link involves a negation of one of the segments) in a way that additive and negative connectives like the adversatives (e.g., even though) would be the most cognitively complex. As far as second language acquisition is concerned, the evidence also suggests an interesting interaction between the type of semantic relationship marked by the connective and its familiarity. Crosson et al. (2008) found that children learning English as a second language present more difficulties when facing temporal, causal or adversative low familiar connectives between sentences than additive and contrastive ones. However, when high familiar connectives were used, the level of comprehension was similar for all semantic types, cancelling or smoothing the effect of cognitive complexity. From the point of view of text simplification, the confirmation of this type of interaction in students with ID would lead to the substitution of low with high familiar connectives.

The type and familiarity of connectives were variables that we did not control for in Experiment 1 (different types of connectives were added to the texts wherever they were appropriate to connect sentences), a circumstance that could be producing a confounding effect in the data. A post hoc correlational analysis

between types of connectives and inferential comprehension in Experiment 1 could help to elucidate whether this hypothesis is certain, but the number of texts and connectives by category (additive, temporal, causal, etc.) is too low to allow this type of analysis.

Therefore, Experiment 2 was designed to compare the effects of familiarity and connective type on text comprehension by means of cloze-type task, the Text Cohesion Task (TCT) (adapted from Crosson et al., 2008).

Experiment 2

Hypotheses

H1. In line with previous studies (Fajardo et al., 2010), we predict that students with ID and theirs pairs in reading comprehension (8–9 years) will have more difficulties in a TCT when using low familiarity connectives than high familiarity connectives, showing a delay more than an inability in the acquisition of semantic and syntactic relations.

H2. Following previous findings with regular readers (e.g., Geva, 2006; Shapiro & Hudson, 1991), we predict that students with ID will have more difficulties in a TCT when inferring low familiar temporal, causal or adversative connectives than additive and contrastive ones.

Method

Participants. Sixteen students with ID (6 males and 10 females with an average age of 20, *SD* = 2.2) and two control groups of students without ID: (1) Chronological age-matched control group: 18 Psychology undergraduate (3 males

and 15 females with an average age of 21, *SD* = 2) and (2) reading level-matched control group: 18 children of third grade of elementary school (9 males and 9 females), average age of 8.34 (*SD* = 0.29) participated in this study.

Base-line measures. ***Experimental group***. The inclusion criteria for participants with ID were the same as in Experiment 1: (a) borderline to mild intellectual disabilities and (b) reading comprehension level equivalent to the third grade of elementary school (ages eight to nine). Fourteen of the 19 students with ID who participated in Experiment 1 participated again in Experiment 2. In addition, there were two new students who were evaluated with the same base-line measures as in Experiment 1, that is, the Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1997) and the ECL1 (De la Cruz, 1999). The average scores of students in each measure were again calculated. The IQ composite standard score for K-BIT (combination of vocabulary and matrices scores) of the group was 67.5 (*SD* = 11, 3rd percentile), which corresponds to mild intellectual disability according to the classification of the DSM-IV-TR Manual (American Psychiatric Association, 2000). The average direct score in the ECL1 was

9.1 (*SD* = 2.4), which corresponds to the 50th percentile for the 3rd grade level.

Chronological age-matched control group. The 18 participants of this group were undergraduate students of Psychology. Reading level and IQ were assumed to be regular since they were undergraduate students, so baseline measures were not applied in this group.

Reading level-matched control group. The 18 participants of this group were 3rd graders (8–9 years) of a regular education school. Reading level was tested through the ECL1 test. The average direct score in the ECL1 was 10.61 (*SD*

= 1.54) which corresponds to the 71st percentile for the 3rd grade level. IQ was assumed to be regular since control group children were in a regular school without special education support.

Experimental task. The Text Cohesion Task (TCT; see Appendix B) used by Crosson et al. (2008) was adapted to Spanish. In this task, students are asked to read sentences that are missing a connective that links ideas between two sentences or two clauses and choose the connective that makes the most contextual sense from three choices. In order to select target connectives, two factors were taken into consideration: class of semantic relations and word familiarity of the connective. Regarding class of semantic relations, we selected the five classes used by Crosson et al. (2008), which typically appear in most taxonomies (Graesser, McNamara, Louwerse, & Cai, 2004; Sanders, Spooren, & Noordman, 1992): additive, temporal, causal, adversative and contrastive.

The connective is considered additive when the two linked segments are equally true, add new information, examples or make a restatement to support a previous argument such as *furthermore* or *for example*. Contrastive connectives such us *on the contrary* or *either* indicate oppositions or exceptions between the two non-causally related segments. Temporal connectives indicate a temporal relation between segments such as *prior to* and *after*. Causal connectives signal that one segment occurs as a result of the other, such as *as* or *given that*. Finally, adversative connectives signal a causal relationship between two segments that is in opposition or contrast, for example, *even though* and *although*.

Connective familiarity was pre-tested in a familiarity test composed of 43 connectives, which pertained to each of the above categories. The pre-test was

first performed by an additional chronological age-matched group. Participants were asked to rank connective familiarity from 1 to 7 (1 = most unfamiliar, 7 = most familiar). The mean word familiarity was 5.71 (SD = 1.13). Those connectives below the 33rd percentile (5.11) were considered low familiar connectives, those between the 33rd and the 66th percentile (6.40) were considered medium familiar connectives and those above the 66th percentile were considered high familiar connectives. Within each semantic relation class, a target connective representing both low and high familiarity was selected, except for temporal connectives where there were no low familiar connectives. Finally, for each one of the nine different target items, three two-clause texts were designed (see examples in Appendix B). Target connectives of low familiarity were only matched with low- or medium-familiarity distractor connectives in the TCT. Target connectives that were medium or high familiarity were matched only with medium- and high-familiarity distractor connectives. In addition, distractor connectives did not make sense in the context of the cloze item. Therefore, the TCT task contained 30 items, that is, two-clause texts in which the cloze was embedded. The percentage of correct answers was used as the dependent variable of this task.

Procedure. In the case of the experimental group, all assessments were administered in two sessions. In the first session, a consent form was firstly signed by participants and then baseline measures were applied in a fixed order. In the second session, the 30 trials of the TCT were individually administered on a desktop computer. The program, designed in Visual Basic 6.0 by the researchers, administered the 30 items in three different orders randomly assigned to participants and registered their answers. The adult control group signed the

consent form and then performed the TCT in the same session. In the case of the 3rd graders, the parents did sign the consent before the experimental session.

Design. A within subject design with type of connective (five levels: additive, contrastive, temporal, causal and adversative) and familiarity of connective (two levels: low and high) as independent variables and percentage of correct answers as the dependent variable was used.

Results

The descriptive results (*M, SD* and *Mdn*) are shown in Table 2. The normality criterion was not satisfied, so non-parametric tests were used.

Table 2.

Experiment 2. Percentage of correct answers in the TCT (*M, SD* and *Mdn*) for each condition of type and frequency of connective.

	Percentage of correct answers in TCT M (*SD, Mdn*)					
	EG		Chrono.CG		Read.CG	
	Low fam.	High fam.	Low fam.	High fam.	Low fam.	High fam.
Ad.	.42 (.26, .33)	.69 (.35, .67)	.87 (.17, 1)	.96 (.11, 1)	.28 (.37, .17)	.46 (.31, .33)
Cont.	.42 (.33, .33)	.79 (.36, 1)	.91 (.15, 1)	1 (0, 1)	.48 (.31, .50)	.69 (.33, .67)
Temp.	.85 (.24, 1)	.73 (.33, .84)	1 (0, 1)	.98 (.08, 1)	.65 (.40, .67)	.52 (.33, .67)
Cau.	.73 (.35, 1)	.50 (.37, .33)	1 (0, 1)	1 (0, 1)	.50 (.33, .33)	.61 (.40, .67)
Adv.	.58 (.31, .67)	.67 (.27, .67)	1 (0, 1)	.93 (.14, 1)	.33 (.40, .17)	.48 (.31, .50)

Note. Ad. = additive. Cont. = contrastive. Temp. = temporal. Cau. = causal. Adv. = adversative. EG = experimental group. Chrono.CG = Chronological age control group. Read.-CG = Reading level-matched control group.

H1 stated that students with ID would have more difficulties in a TCT when using low familiarity connectives than high familiarity connectives. In effect, the Wilcoxon

matched pairs test confirmed that, for the experimental group, the percentage of correct answers in the TCT was higher in the high familiarity condition than in the low familiarity condition with the type of connective collapsed (Z = 2.02; p = .042). Neither in the case of the adult control group nor in the children's the effect of connective familiarity was significant (Z = 1.86; p = .063, Z = 1.90; p = .06). The Kruskal–Wallis analysis revealed significant differences among groups in both high (H(2) = 31.39; p = .00) and low familiarity (H(2) = 31.45; p = .00) so Mann– Whitney U analysis were performed to explore those differences. Mann–Whitney U tests showed that adult control students obtained significantly higher scores in the TCT than participants with ID in both high (Z = -4.11; U = 25; p = .00) and low familiarity items (Z = -4.28; U = 20; p < 001). They also performed significantly better than the children's group in high (Z = -5.13; U = 0; p = .00) and low familiarity connectives (Z = 5.06; U = 2; p = .00). The comparison between the group of students with ID and the 3rd graders was not significant neither in high (Z = 1.45; U = 102; p = .15) nor in low (Z = 1.36; U = 104; p = .17) familiarity conditions.

The familiarity facilitation hypothesis (H1) was tested again for each type of connective and group by means of the Wilcoxon matched pairs test. In the experimental group, the advantage of high familiarity conditions was only observed for the two types, additive, Z = 2.5; p = .013, and contrastive, Z = 2.93; p = .003. However, they obtained higher scores in the TCT with temporal and causal connectives of low familiarity than the same type of connectives with high familiarity (temporal, Z = 2.2; p = .027; causal, Z = 2.13; p = .033). In the case of the 3rd graders, the familiarity facilitation hypothesis was confirmed for additive (Z = 2.43; p < .02) and contrastive connectives (Z = 1.89; p < .059, marginally significant difference in this case).

The adults' control group, whose percentage of correct answers was above 87% in all conditions, only took advantage of high familiarity in contrastive connectives ($Z = 2.02$; $p = .04$).

H2 stated that students with ID will have more difficulties in a TCT when inferring low familiar temporal, causal or adversative connectives than additive and contrastive ones. Therefore, in order to test this hypothesis, we analysed the effect of ''semantic class of the connector'' for each level of familiarity. For the group of students with ID, Wilcoxon matched pairs tests confirmed an effect of type of connective (class of semantic relation) but not in the predicted way, especially in the case of low familiar connectives (see Table 2). For high familiar connectives, students with ID obtained higher scores in the TCT with contrastive connectives, followed by temporal connectives. They obtained lower scores with causal connectives, followed by adversative connectives, but the difference was only significant between contrastive and causal conditions ($Z = 2.67$; $p < .01$) and causal and temporal ($Z = 2.24$; $p = .02$). However, for low familiar connectives, students with ID obtained a higher percentage of correct answers in the TCT with temporal and causal connectives followed by adversative, additive and contrastive (results resumed in Table 2), that is, just the opposite pattern to what was predicted (more difficulties in a TCT when using temporal, causal or adversative connectives than additive and contrastive ones). In the high familiarity condition, the reading level-matched control group obtained the highest scores in contrastive and causal followed by temporal, adversative and additive connectives. Only the differences between contrastive and additive were significant ($Z = 2.09$; $p = .04$). Regarding low familiarity connectives, the reading level-matched control group obtain the highest scores with temporal connectives, followed by causal, contrastive, adversative and additive. The difference between temporal and additive and

adversative connectives was significant (temporal vs. additive, $Z = 2.67$; $p = .01$; temporal vs. adversative, $Z = 2.23$; $p = .03$). That is, opposite to the predicted pattern, low familiarity additive connectives were the hardest to infer by the reading level-matched control group.

Discussion

In Experiment 2, two hypotheses were tested. H1 predicted a facilitative effect of connective familiarity that was supported in general terms, that is, all groups obtained higher general performance in the TCT when using high familiar connectives than low familiar connectives. However, there was an interaction of familiarity with type of connectives, which constrained the facilitation of familiarity to additive, contrastive and, showing a non-significant trend, adversative connectives, while low familiar temporal (*prior to*) and causal (*given that*) connectives were easier to infer by readers with ID than high familiar temporal (*after that*) and causal (*as*) connectives. In the case of causal connectives, a plausible explanation for this effect is that the high familiar causal connective used in this experiment (*as*) also has more senses than the low familiar connective (i.e., as can be temporal, causal or adversative while given that can only be causal; the same happens with the original Spanish conjunctions *pues* vs. *puesto que*). Therefore, the process of decision making between target and distractors in the TCT would have been more difficult for the high familiar connective [in spite of the fact that, following Crosson et al. (2008), distractors were carefully selected to ensure that they did not make sense in the context of the cloze item. That is, as noted before by Crossley et al. (2007), the fact that highly familiar words in languages such as English or Spanish tend to be polysemous could mean that

texts that contain them become ambiguous as a whole, so they would be an inadequate simplification solution.

In the case of temporal connectives, a methodological flaw could explain the facilitation of the low familiar connective (*prior to*). As we noted in the Method section, the low familiar temporal connective was actually of medium familiarity since no low familiarity connective for this class was obtained in the familiarity test. This fact could explain why the temporal connective was significantly easier to infer in the TCT than the rest of the low familiar connectives. Therefore, in a strict sense, only the rest of the connectives should be considered in our analysis of the type of connective effect.

The interaction between familiarity and type of connective also qualified H2, which – based on previous findings (e.g., Geva, 2006; Shapiro & Hudson, 1991) and according to the cognitive complexity theory (Spooren & Sanders, 2008) – stated that the order of connectives from easiest to most difficult to process would be as follows: additive, contrastive, temporal, causal, and adversative. However, additive and contrastive connectives proved easier to process only in the high familiar condition, while the opposite pattern was observed in the low familiar condition where temporal connectives (followed by causal and adversative connectives) were easier to infer by readers with ID than additive and contrastive connectives. Actually, only 40% of answers in TCT trials with low familiar additive and contrastive connectives were correct, a lower rate overall for students with ID.

The cognitive complexity theory was neither supported for the 3rd graders group who, as the group of students with ID, obtained the highest scores in the TCT for the low familiarity temporal connective condition. However, the facilitation of low familiarity causal connectives was not found in this group. A possible explanation of this result is that the polysemy of the causal connective could be

overcome by this group of students using a more general decision making strategy in the TCT like discarding the connective options which did not make sense and selecting the remaining connective even if they were not sure about its meaning, strategy that could be unavailable for the group of students with ID.

Another interesting result concerns the comparison between groups. In general terms, as the 3rd graders control group, the poor readers with ID were less able to select the appropriate answer in TCT than the chronological-age control group. This result agrees with those observed recently in younger readers (10- and 8-year olds) by Cain and Nash (2011). However, the different profiles of connective facilitation between the ID group and the 3rd graders control group suggests that readers with ID present a specific deficit more than a simple delay in the acquisition of connectives.

Finally, it is worthy to discuss the lack of support to the cognitive complexity theory in the case of the 3rd graders control group. This result could be explained by the fact that the particular connectives used in our experiment for each semantic class was different to the used in other experiments. As highlighted by Cain and Nash (2011), different connectives of a particular class of semantic relation between segments might indicate different types of temporality, causality or addition being more difficult, for example, to infer *for example* than *and* although they both pertain to the category of ‘‘additive’’ connectives.

Conclusions

Low reading literacy achievement of students with ID is noticeably constraining their academic success and consequently their access to information, employment opportunities and even their entertainment options. This research was

conducted in order to meet the educational challenge identified by Morgan and Moni (2008) of overcoming the limited availability of ‘‘literacy resources’’ for adolescents and adults with ID by testing the effects of some linguistic variables that could be deliberately measured and manipulated to select and design suitable texts.

In particular, we focused on connectives (e.g., but, in contrast, etc.), cohesive elements of texts, which are expected to facilitate inferential comprehension in poor readers. This prediction was tested by means of two experiments. Experiment 1 compared literal and inferential text comprehension of texts with and without connectives and/or high frequency content words. That is, we compared a deep readability measure (connective presence), expected to have a higher impact in inferential comprehension with a shallow-based measure (content word frequency), expected to affect literal comprehension (Crossley et al., 2007). The results indicated that neither the addition of high frequency content words nor connectives in general produced inferential comprehension improvements. A possible explanation for connectives' lack of facilitation is that such facilitation would depend on the class of semantics relation made explicit by the connective. This hypothesis was then tested in Experiment 2. In addition, this second experiment explored the effect of a second variable, the connective familiarity and its interaction with connective type on two-clause text comprehension by means of a cloze task (Text Cohesion Task – TCT). In particular, we predicted that high familiar connectives would be easier to infer in the TCT than low familiar connectives and, according to the cognitive complexity theory (Spooren & Sanders, 2008) and previous findings from the language acquisition field, we hypothesised that additive and contrastive connectives would

be more easily inferred than temporal, causal and adversative connectives, especially when familiarity was low.

The results of Experiment 2 showed that, although readers with ID were less likely to select the target connective in the cloze task than chronological age-matched readers in general and similar, also in general terms, to reading age-matched, their performance level was affected by the type of connective and its familiarity. They obtained lower comprehension scores with low familiar additive connectives (e.g., furthermore) and contrastive connectives (e.g., on the contrary) and higher comprehension scores with low familiar temporal (e.g., prior to) and causal (given that) connectives. One of the most plausible explanations for this interaction effect between familiarity and type of connective is the polysemy of high familiar words in languages like English or Spanish (see discussion in Crossley et al., 2007). For instance, high familiar connectives also have more senses than low familiar connectives (e.g., as can be temporal, causal or adversative, while given that can only be causal), so the process of decision making between target and distractors in the TCT would have been more difficult for high familiar connectives. Reading level-matched control group with regular intellectual abilities, who did not show this reverse effect of familiarity for causal connectives, could have overcome the polysemy problem by applying a more general decision making strategy in the cloze task like discarding those connectives that do not make sense and select the remaining item even if you were not sure about its meaning. That conclusion highlights that the cloze task used in this experiment is not just a reading comprehension task but a decision making one so both processes, reading comprehension and decision making, could have been confound in this study. Future research should overcome this limitation by measuring reading comprehension by means of open comprehension questions.

Alternatively, as noted before by Cain and Nash (2011), within the group of connectives that express a particular class of semantic relation between segments, different connectives might indicate different types of temporality, causality or contrast. Therefore, instead of paying attention to the theoretical connective semantic category, it may be more important to consider specific connectives because they could differ in cognitive complexity regardless of their category.

The hypothesis of the ‘‘particular connective’’ could also explain the results of the children control group who, although obtain a similar low level of performance in general in the TCT than the experimental group, its pattern of results for each type of connective and familiarity varied and was contrary to the found in previous studies.

From an applied point of view, our results shed some light on how to simplify and select suitable texts for adolescents and adults with ID, which is one of the challenges of the official educational programs like the one framing this research, the ‘‘Lectura Fácil’’ project (funded by the Spanish Ministry of Industry in 2009), aimed to promote newspaper reading in students with ID. In particular, as Cain and Nash (2011) recommend for children with poor reading skills: ‘‘[. . .] in contrast to the conclusions of text readability formulae, a longer sentence in which the link between two clauses is explicitly signalled may be easier to understand than two short separate sentences, if the individual has reached a certain level of knowledge of the specific connective. Thus, appropriate and informative connectives appear to help, not hinder, young readers to process and understand written text. Connectives, as well as other markers of text cohesion, should inform calculations of text complexity to identify suitable texts for different levels of reader’’ (p. 439). Our findings suggest that this conclusion could also be applied to poor readers with

ID. Future research should investigate in depth which are the specific connectives that facilitate text segment integration for young students and adults with mild ID.

Finally, it is important to highlight methodological limitations of our research. The first one is that, in spite of using within-subject designs, the sample sizes of our two experiments are not large, so the lack of effects in Experiment 1, for example, would have been due to a lack of statistical power. The second one concerns the genre of the journalistic texts used in Experiment 1 that were narrative, expository or a mixture of both. As showed by Bos and Tierney (1984), text cohesion is easier to establish in narrative texts than in expository ones by readers with ID. Therefore, a potential way of simplify journalist texts could be change the genre from expository to narrative when possible. Unfortunately, we did not control for this variable in our study so future research should address this issue explicitly. The third limitation refers to the specific connectives used in Experiment 2, which were selected based on a pilot study with students without ID, a fact that could be an alternative explanation to the polysemy explanation for the paradoxical detrimental effects of some high familiar connectives in the sample of readers with ID and the 3rd graders control group. The fourth limitation concerns the use of the cloze task in Experiment 2, which is not simply a test of reading comprehension but involves also decision making which might have make the task even more difficult for reading with ID. At least, these four particular methodological drawbacks should be overcome in future research.

References

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: contrasting frequency effects in Spanish. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. http://doi.org/10.1037/0278-7393.27.2.545

American Psychiatric Association. (2000). *DSM-IV*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Association.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *CELEX Lexical Database Release 2*. Philadelphia: Linguistics Data Consortium.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language*, *7*(2), 235–261. http://doi.org/10.1017/S0305000900002610

Bos, C. S., & Tierney, R. J. (1984). Inferential reading abilities of mildly mentally retarded and nonretarded students. *American Journal of Mental Deficiency*, 89(1), 75–82.

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. http://doi.org/10.1037/a0022824

Cain, K., Patson, N., & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, *32*(04), 877. http://doi.org/10.1017/S0305000905007014

Conners, F. A. (2003). Reading skills and cognitive abilities of individuals with mental retardation. In Abbeduto (Ed.), *International review of research in mental retardation* (27th ed., pp. 191–229). New York: Academic Press.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & Mcnamara, D. S. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, *91*(1), 15–30. http://doi.org/10.1111/j.1540-4781.2007.00507.x

Crosson, A. C., Lesaux, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, *29*(04), 603–625. http://doi.org/10.1017/S0142716408080260

Deane, P., Sheehan, K. M., Sabatini, J., Futagi, Y., & Kostin, I. (2006). Differences in text structures and its implications for assessment of struggling readers. *Scientific Studies of Reading*, 10, 257–275.

De la Cruz, M. (1999). *Evaluación de la Comprensión Lectora (ECL-1)*. Madrid: TEA.

Doctorow, M., Wittrock, M. C., & Marks, C. (1974). Generative processes in reading comprehension. *Journal of Educational Psychology*. http://doi.org/10.1037/0022-0663.70.2.109

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III: Peabody test de vocabulario en imágenes*. Madrid: TEA.

Fajardo, I., A´ vila, V., Tavares, G., & Ferrer, A. (2010). Easy to read text for students with intellectual disability. In C. Ma¨hler (Chair), *Preschool intervention studies and improving reading*. Paper presented at the Biennial Meeting of EARLI SIG 15: Special Educational Needs, Frankfurt am Main, Germany.

Fernández-Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna*, *214*, 29–32.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, *32*(3), 221–233.

Freebody, P., & Anderson, R. C. (1983). Effects of Vocabulary Difficulty, Text Cohesion, and Schema Availability on Reading Comprehension. *Reading Research Quarterly*, *18*(3), 277–294. http://doi.org/10.2307/747389

Geva, E. (2006). Conjunction use in school children's oral language and reading. In R. Horowitz (Ed.), *Talking texts: How speech and writing interact in school learning*. Hillsdale, NJ: Erlbaum.

Goodman, K. & Freeman, D. (1993). What's simple in simplified language. In: M.L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 69–76). Singapore: SEAMEO Regional Language Center.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(2), 193–202. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15354684

Karreman, J., Van Der Geest, T., & Buursink, E. (2007). Accessible website content guidelines for users with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *20*, 510–518. http://doi.org/10.1111/j.1468-3148.2006.00353.x

Kaufman, A. S., & Kaufman, N. L. (1997). *Test breve de inteligencia de Kaufman (K-BIT)*. Madrid: TEA.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. http://doi.org/10.1037/0033-295X.95.2.163

Long, M., & Ross, S. (1993). Modifications That Preserve Language and Content. *Simplification: Theory and Application. Anthology Series 31*, 29–52. Retrieved from http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED371576

Mastropieri, M. a., & Scruggs, T. E. (1992). Science for Students With Disabilities. *Review of Educational Research*. http://doi.org/10.3102/00346543062004377

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes*. http://doi.org/10.1080/01638530902959943

Millis, K. K., & Just, M. A. (1994). The Influence of Connectives on Sentence Comprehension. *Journal of Memory and Language*, *33*(1), 128–147. http://doi.org/10.1006/jmla.1994.1007

Moni, K. B., & Jobling, A. (2001). Reading-related Literacy Learning of Young Adults with Down Syndrome: Findings from a three year teaching and

research program. *International Journal of Disability, Development and Education*, *48*(4), 377–394. http://doi.org/10.1080/10349120120094275

Morgan, M. F., & Moni, K. B. (2008). Meeting the challenge of limited literacy resources for adolescents and adults with intellectual disabilities. *British Journal of Special Education*, *35*(2), 92–101. http://doi.org/10.1111/j.1467-8578.2008.00378.x

Nash, H., & Heath, J. (2011). The role of vocabulary, working memory and inference making ability in reading comprehension in Down syndrome. *Research in Developmental Disabilities*, *32*(5), 1782–1791. http://doi.org/10.1016/j.ridd.2011.03.007

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*(5), 573–590. http://doi.org/10.1016/0749-596X(92)90029-W

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: the passage or the question? *Behavior Research Methods*, *40*(4), 1001–1015. http://doi.org/10.3758/BRM.40.4.1001

Pany, D., Jenkins, J. R., & Schrenk, J. (1982). Vocabulary instruction: Effects on word knowledge and reading comprehension. *Learning Disability Quarterly*, *5*, 202–215. http://doi.org/10.2307/1510288

Ramos, J. L., & Cuetos, F. (1999). *PROLEC-SE: evaluación de los procesos lectores en alumnos del tercer ciclo de educación primaria y educación secundaria obligatoria*. Madrid: TEA.

Ryder, R., & Hughes, M. (1985). The effect on text comprehension of word frequency. *Journal of Educational Research*, 78, 286–291.

Sanders, T., Land, J., & Mulder, G. (2007). Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal*, 15(3), 219–235.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. http://doi.org/10.1080/01638539209544800

Shapiro, L. R., & Hudson, J. a. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, *27*(6), 960–974. http://doi.org/10.1037/0012-1649.27.6.960

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics*, *40*(12), 2003–2026. http://doi.org/10.1016/j.pragma.2008.04.021

Tronbacke, B. (1997). *Guidelines for Easy-to-read Materials*. (I. Headquarters, Ed.).

Verhoeven, L., & Vermeer, A. (2006). Literacy achievement of children with intellectual disabilities and differing linguistic backgrounds. *Journal of Intellectual Disability Research : JIDR*, *50*(Pt 10), 725–38. http://doi.org/10.1111/j.1365-2788.2006.00838.x

Appendix A

Extract of one of the journalistic texts used in Experiment 1 translated into English (the original texts were in Spanish) in three versions: non-adapted version, frequency adapted and connectivity adapted. Note that the order of words for Spanish and English versions varied due to syntactic differences between the languages. One example of the two types of comprehension questions that accompanied the text, literal and inferential, are also provided.

Non-adapted version (Spanish)

El ministro de Trabajo, Celestino Corbacho, ha reiterado (1) que el paro "va a seguir creciendo" en los próximos meses después de que en agosto se rompiera una racha (2) de tres meses consecutivos (3) de caída del desempleo (4). Según Corbacho, históricamente los meses de otoño han registrado (5) subidas del desempleo (6). Como ha afirmado, "No hay por qué pensar que este trimestre tiene que ser diferente". Ha afirmado que el Gobierno prevé (7) que este trimestre "no va a tener la dureza" del mismo en el año pasado.

Non-adapted version (English)

Minister of Labour, Celestino Corbacho, has insisted (1) that the unemployment rate will keep rising during the next months after the idleness (4) propensity (2) to rise was stopped during three consecutive (3) months starting in August. According to Corbacho, autumn months have historically registered (5) idleness (6) risings. As he affirmed, ‘‘There is no reason to think that this trimester has to be different’’. He has affirmed that the government anticipates (7) that this trimester is not going to be as hard as last year’s trimester.

Frequency adapted version (Spanish)

El ministro de Trabajo, Celestino Corbacho, ha dicho que el paro "va a seguir creciendo" en los próximos meses después de que en agosto se rompiera una temporada de tres meses seguidos de caída del paro. Según Corbacho, históricamente los meses de otoño han tenido subidas del paro. Como ha afirmado, "No hay por qué pensar que este trimestre tiene que ser diferente. Ha afirmado que el Gobierno cree que este trimestre "no va a tener la dureza" del mismo en el año pasado.

Frequency adapted version (English)

Minister of Labour, Celestino Corbacho, has said (1) that the unemployment rate will keep rising during the next months after the unemployment (4) tendency (2) to rise was stopped during three repeated (3) months starting in August. According to Corbacho, the autumn months have historically shown (5) unemployment (6) risings. As he affirmed, ‘‘There is no reason to think that this trimester has to be different’’. He has affirmed that the government expects (7) that this trimester is not going to be as hard as last year’s trimester.

Connectivity adapted version (Spanish)

El ministro de Trabajo, Celestino Corbacho, ha reiterado que el paro "va a seguir creciendo" en los próximos meses después de que en agosto se rompiera una racha de tres meses consecutivos de caída del desempleo. Según Corbacho, históricamente los meses de otoño han registrado subidas del desempleo. Por tanto (1), como ha afirmado, "No hay por qué pensar que este trimestre tiene que ser diferente”. Sin embargo (2), ha afirmado que el Gobierno prevé que este trimestre "no va a tener la dureza" del mismo en el año pasado.

Connectivity adapted version (English)

Minister of Labour, Celestino Corbacho, has insisted that the unemployment rate will keep rising during the next months after the idleness propensity to rise was stopped during three consecutive months starting in August. According to Corbacho, the autumn months have historically registered idleness risings. Thus (1), as he affirmed, ‘‘There is no reason to think that this trimester has to be different’’. However (2), he has affirmed that the government anticipates that this trimester is not going to be as hard as last year’s trimester.

Comprehension questions (Spanish):

3. ¿Qué noticia ha dado el ministro de trabajo? **Literal**
   d. Que el paro va a seguir creciendo. **Correct**
   e. Que va a haber menos paro que el año pasado.
   f. Que va a haber el mismo paro que el año pasado.

4. ¿Por qué se espera una subida del paro este otoño? **Inferential**
   d. Porque lo dice el ministro Corbacho.
   e. Porque muchas personas estarán en paro.
   f. Porque todos los años sube el paro en otoño. **Correct**

Comprehension questions (English):

1. ¿Which news has the Minister of Labour announced? **Literal**
   a. The unemployment rate will keep rising. **Correct**
   b. The unemployment rate will be lower than last year.
   c. The unemployment rate will be the same as last year.

2. ¿Why is a rise in unemployment expected this autumn? **Inferential**
   a. Because Minister Corbacho has said so.
   b. Because many people will be unemployed.
   c. Because the unemployment rate rises every autumn. **Correct**

Appendix B

Example of the Text Cohesion Task (TCT) adapted from Crosson et al. (2008). Our original version was in Spanish.

Class of semantic relation	Familiarity	Target Connective	Distractor	Distractor	Example sentence
Ad.	Low	asimismo	verbigracia	tal como	A Tomás le gusta jugar a la videoconsola. Asimismo le gusta pintar.
		furthermore			*Tomás likes to play videogames. Furthermore, he likes to paint.*
Ad.	High	por ejemplo	incluso	así que	Julia odia los colores vivos, por ejemplo el amarillo, el rosa y el verde.
		for example			*Julia hates bright colours, for example, yellow, pink and green.*
Cont.	Low	por el contrario	por lo tanto	pese a	El animal preferido de Susana es el perro. Por el contrario, Carolina prefiere los gatos.
		on the contrary			*The favourite animal of Susana is the dog. On the contrary, Carolina prefers cats.*
Cont.	High	tampoco	luego	ya que	Mario sólo habla español. Pedro tampoco habla otros idiomas.
		either			*Mario only speaks Spanish. Pedro cannot speak other languages either.*
Temp.	Mid	antes de (que)	de modo que	por consiguiente	Mi abuela nació antes de que terminara la Guerra Civil Española.
		prior to			*My grandmother was born prior to the end of the Spanish Civil War.*
Temp.	High	después	además	cuando	Estudiaron para el examen más de cuatro horas. Después fueron a pasear para relajarse.
		After that			*They studied for the exam for more than four hours. After that, they went for a walk to relax.*

Cau.	Low	puesto que	verbigracia	tal como	Necesita ganar más dinero, puesto que tiene que mantener a su familia.
		given that			*He needs to earn more money given that he has to support his family.*
Cau.	High	pues	también	cuando	Mañana haré un regalo a mi madre, pues es su cumpleaños.
		as			*Tomorrow, I'll give a present to my mother as it is her birthday.*
Adv.	Low	pese a	ya que	de otra forma	Sonia piensa que María es su mejor amiga pese a que se pelean continuamente.
		even though			*Sonia thinks that María is her best friend, even though they argue frequently.*
Adv.	High	aunque	además	sin embargo	Jaime no estaba contento aunque su equipo había ganado el partido de baloncesto.
		although			*Jaime was not happy, although his team had won the basketball match.*

Note. Ad. = additive. Cont. = contrastive. Temp. = temporal. Cau. = causal. Adv. = adversative.

5 Study 3

Who do you refer to? How young students with mild intellectual disability confront anaphoric ambiguities in texts and sentences.

Tavares, G., Fajardo, I., Ávila, V., Salmerón, L., & Ferrer, A. (2015). Who do you refer to? How young students with mild intellectual disability confront anaphoric ambiguities in texts and sentences. *Research in Developmental Disabilities: RIDD*, 38, 108–124. http://doi.org/10.1016/j.ridd.2014.12.014

Who do you refer to? How young students with mild intellectual disability confront anaphoric ambiguities in texts and sentences

Gema Tavares*, Inmaculada Fajardo, Vicenta Ávila, Ladislao Salmerón, Antonio Ferrer

*ERI Lectura, University of Valencia

Abstract

Along 2 experiments we tested the anaphoric pronoun resolution abilities of readers with intellectual disability in comparison with chronological and reading age-matched groups. In Experiment 1, the anaphor test of Elosúa, Carriedo and García-Madruga (2009) confirmed that readers with intellectual disability (ID) are slower than control readers resolving clitic anaphoric pronouns, especially when the use of morphological cues (e.g. gender) is necessary. In order to test if the poor performance could be due to low levels of metacognitive skills during reading, an inconsistency detection task combined with eye tracking was designed in Experiment 2. Participants read short texts with an anaphoric pronoun in the fifth sentence, either morphologically (gender) consistent or not with the information provided in the second sentence. The scores in the anaphor comprehension questions presented after the text confirmed that readers with ID are affected by the gender inconsistency but they are unable to explicitly report it and recover from it, as the number of re-fixations after reading the critical sentence suggests. As their answers to the explicit detection questions showed, the adults control group did not show any preference for morphosyntax or semantics in spite of being aware of the inconsistency. In sum, both groups of readers with and without ID are affected by inconsistencies, but ID readers do not have appropriate metacognitive skills to explicitly identify the source of the inconsistency and fix it.

Keywords: anaphora resolution; inconsistency detection; intellectual disabilities; metacognitive skills; text comprehension

Who do You Refer to? How Young Students with Mild Intellectual Disability Confront Anaphoric Ambiguities in Texts and Sentences

Literature review

An anaphor is a linguistic reference to an antecedent piece of text (Rademaker & Haeusler, 2008). Though the form of an anaphor is varied: repetition (e.g. *the cake* and the coffee are in the table. *The cake* is delicious.), pronouns (*the cake* - *it*), lexical co-reference (*the cake* - *the dessert*) or even ellipsis (*the cake* - Ø), its function is always to maintain text cohesion by recalling previously mentioned entities (antecedents) without the need of merely repeating them. Struggling readers such as readers with Intellectual Disability (ID), who are the target of the present study may lack the adequate skills to identify and understand an anaphor, which is core for text comprehension.

How an anaphor is resolved depends on intralinguistic features that is, related to the text itself: word length (e.g. number of characters or syllables), word frequency of the anaphor antecedent (e.g. *domicile* vs. *house* as possible antecedents of the anaphoric pronoun *it*), distance between the antecedent and the anaphor (e.g. number of sentences or words between them) and kind of anaphor (e.g. pronoun, repeated name or ellipsis) (Arnold, Brown-Schmidt, & Trueswell, 2007; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Cacciari, Carreiras, & Barbolini, 1997; Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981; Garvey, Caramazza, & Yates, 1975; Gelormini-Lezama & Almor, 2011; Gordon, Grosz, & Gilliom, 1993; Järvikivi, van Gompel, Hyönä, & Bertram, 2005) as well as extra linguistic factors (related to the reader), mainly working memory (the memory system in charge of temporarily storing and managing the

information, Baddeley, 1992) and metacognitive skills such as planning, checking and revising strategies during reading comprehension (Cain, Oakhill, Barnes, & Bryant, 2001; Cain & Oakhill, 1999; Ehrlich, Remond, & Tardieu, 1999; Long & De Ley, 2000; Yuill & Oakhill, 1988) .

Apart from the semantics of the antecedent, in some languages such as Spanish pronominal anaphors bear a grammatical load for gender and number that helps to establish the link with the antecedent. Thus, finding a suitable antecedent for a pronoun can be accomplished through either (or both) of these grammatical features. Imagine a sentence like "Peter (m) gave one more coin to Thomas (m). He (m) had too many". Here, both *Peter* and *Thomas* could be the referred antecedent of the pronoun *he*. However, the context indicates that the most likely situation would be Peter giving out a coin as the result of having too many. In "Maria (f) gave one more coin to Thomas (m).He (m) had too many" *Maria* is the one awarding *Thomas* with a coin. A first probabilistic analysis will lead us to the same conclusion as in the former example. Nonetheless, the gender load of the pronoun indicates that the antecedent should be a masculine entity, becoming Thomas the only possible candidate. As one might expect, this duality is sometimes problematic resulting in the reader finding troubles to attach the pronoun to its proper antecedent.

There exist some antecedents on this psycholinguistic issue, as for instance the study of Oakhill and Yuill (1986) that explored the inferences drawn during a pronoun resolution task performed by skilled and low-skilled 7-8 year olds, who read two-clause subordinate sentences where the proper names were either of the same or different gender (e.g. Peter lent ten pence to Liz because she was very poor). In a first experiment, the subordinate clause was introduced by a pronoun referring to either the subject or the object of the main clause; in the

second one, a gap was presented instead of the pronoun for participants to fill it in. Additionally, a comprehension question was presented right after the stimulus in the first experiment. Results showed that low-skilled readers encountered more difficulties drawing inferences about pronominal antecedents than their skilled peers, even when there was a gender clue to link the pronoun to the correct antecedent. These difficulties appeared especially when inferences were complex or included a higher memory load (i.e. the proper names were not given again in the question, so the reader should remember them). The authors found no interaction between level of reading comprehension and memory load, as both groups of readers performed better in the simpler conditions, always with lower accuracy rates in the low-skilled readers. An explanation in terms of metacognitive skills is suggested: low-skilled readers sometimes decide not to go back to the disambiguating information trusting thus in their representation of the text, even though sometimes this is not accurate. Skilled readers, on the contrary, go back in the text when they have doubts about their own representation. In addition, it was also argued that low-skilled readers did not pay enough attention to cues such as the gender of the antecedent.

Megherbi and Ehrlich (2005) also corroborated the conflict of ambiguous pronouns and inference making in spoken language. In this case they followed the hypothesis that 7-8 year children presenting problems in pronoun resolution by reading, should also struggle with them in spoken language. In an anaphoric resolution task children had to decide whether the final word of a sentence should be a masculine or feminine pronoun (e.g. according to the fairy tale, Cinderella put on a beautiful dress to meet the handsome prince. She danced with… him / her). Their findings showed a higher influence of the verb bias (tendency to link the pronoun to the subject or the object) on low-skilled readers, meaning that they

trusted the semantics of the verb more than the gender of the pronoun itself in order to disambiguate it. On the other hand, in line with Oakhill and Yuill (1986)'s findings skilled comprehenders took advantage of the gender cues of the pronoun over verb bias, allowing them to get an extra benefit when there existed no conflict between both linguistic features.

The results of these two studies agree with the findings of Elosúa, Carriedo and García-Madruga (2009). Elosúa et al. (2009) investigated the resolution of clitic pronominal anaphora when morphosyntax (e.g. gender and number clues) and semantics come to conflict. Clitics are a kind of pronouns devoted to act as a direct or indirect object, in the case of Spanish from either a pre-verb or a post-verb position (e.g. pre-verb position: Lisa *lo* cogió mientras *lo* miraba; Lisa took it while staring at it; post-verb position: La madre gritó: "¡Lisa, cóge*lo*!"; the mother shouted: "Lisa, take it!"). In their experiment 1b, Elosúa et al. (2009) compared the performance, in terms of speed and accuracy, of undergraduates in an anaphora resolution task in which the anaphor could be either morphosyntactically or semantically resolved. Results indicated that participants performed equally well in the two types of disambiguation, morphosyntactic and semantic, though it took significantly longer to resolve the semantic ambiguity than the morphosyntactic one. The authors also found a facilitative effect of morphosyntax over semantics, very likely due to the nature of the analysis required to benefit from them and the cognitive effort needed: using a semantic strategy implies the performance of a deeper and more resource-demanding analysis, while following the morphosyntactic clues only requires a surface analysis of the lexical units and the grammar rules.

These antecedents explored anaphora resolution by high- and low-skilled readers but not by youngsters with intellectual disabilities. Though the literature

on this matter and population is scarce, a brief revision of the antecedents is made in the next paragraphs as we consider that assuming that readers with ID are low-skilled readers is a fair simply and inaccurate assumption.

Our own previous research confirmed that young readers with ID experience problems when they have to use and comprehend cohesive elements like connectives (e.g. but, besides, for that reason) in text. More specifically, Fajardo, Tavares, Ávila, and Ferrer (2013) found that readers with ID were less likely to select the target connective in a cloze task than chronologically age-matched readers (mean age = 21 years). Conclusions by Fajardo et al. (2014) added that the number of co-references (anaphors) contained in journalistic texts predicted negatively the literal comprehension (Fajardo et al., 2014). In other words, the higher the number of co-references (whatever the type) the lower the scores in literal comprehension questions by poor readers with ID.

To our knowledge, the only previous existing literature regarding the processing of pronouns was conducted with individuals with a specific aetiology related to ID, such as Down syndrome (DS). For instance, Perovic (2006) focused on how DS individuals (17-21 years) resolved reflexive pronouns in comparison with English typically developing children (5-11 years). Though participants from both groups were matched in receptive vocabulary, DS participants presented problems in the resolution of reflexives (herself) but not in pronouns (her), following the opposite pattern of acquisition to Typically Developed (TD) children. In other words, the acquisition of such co-referentials by DS was not delayed but qualitatively different, since previous research (Jakubowicz, 1991; Santiuste, 1997) had argued that reflexives in TD children are acquired earlier than pronouns.

In sum, previous research has evidenced that individuals with ID could present singular patterns of pronoun acquisition in comparison with regular

readers. In addition, they might find difficulties using cohesive elements such as connectives and co-references during reading, though there are no antecedents in the literature testing how anaphors, and particularly pronominal anaphors, are resolved by this population. In order to overcome this lack of research we designed Experiment 1, where two groups of young readers, one conformed by ID students and the other by TD undergraduates were asked to perform the abovementioned anaphor task of Elosua et al. (2009). The methodological details and operative hypotheses of this experiment are exposed in the next section. In order to test if the observed differences between groups in Experiment 1 could be due to metacognitive factors during reading, an inconsistency detection task combined with eye tracking was designed in Experiment 2.

Experiment 1

Hypotheses

The main goal of the present study was to obtain comprehensive data about how students with unspecific intellectual disabilities resolve clitic anaphors by measuring reaction times and accuracy in an anaphor decision task.

H1a. Overall accuracy is expected to reach lower levels in both conditions, semantic and morphosyntactic, in the experimental group (readers with unspecific intellectual disabilities) than in the control group. In fact, a ceiling effect is expected for the later according to the results obtained by Elosúa et al., (2009).

H1b. Accordingly, overall reaction times should be higher in the experimental group than in the control group.

H2. Finally, given some of the antecedents reviewed on low-skilled readers, (Megherbi & Ehrlich, 2005; Yuill & Oakhill, 1988) the experimental group is expected to show a preference for semantics over morphosyntax reflected by increased times and lower accuracy in this latter condition. Following Elosúa cf. (2009), the opposite patter regarding response times is expected for the control group, that is, faster response times in the morphosyntactic than in the semantic condition.

Method

Participants. ***Experimental group***. Twenty-nine students with ID (15 males and 14 females with an average age of 19.72, *SD* = 2.2, ranging from 16 to 24) recruited at *Camí Obert* (Valencian Community, Spain) and participated voluntarily in the study. Camí Obert is a Vocational Training Centre for People with Special Needs (Associated to APSA, an Association for People with Intellectual Disability of Alicante, Spain). According to the PCPI (the Spanish initials for the First Professional Qualification Programme) Regulating Ordinance released in 2008 (Valencian Community), this kind of teaching programs last 2 years and only admits young people (aged from 16 to 21 years) with an official accreditation of disability who has fulfilled the period of mandatory scholarship (10 years in Spain) and with enough personal and social autonomy as to allow them to participate in the learning process as well as to access and maintain a job. The vehicular language of Camí Obert is Spanish.

The inclusion criteria for participants with ID were:

(a) Borderline to moderate intellectual disabilities (Consequently, students with regular IQ or severe degree of intellectual disability were excluded). The Spanish standardization of The Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1997) was used to measure the verbal and non-verbal intelligence of the students. This is a brief, individually administered measure of the intelligence of a wide range of individuals spanning the ages of 4–90 years. It is composed of two subtests: Vocabulary (Expressive Vocabulary and Definitions) and Matrices. It takes 15–30 minutes to administer. It was administered by the educational psychologists of the centre *Camí Obert* since it was part of the assessment protocol of the centre. The IQ Composite (a combination of vocabulary and matrices scores) internal consistency coefficient was .98 across ages (.88 to .96). The IQ composite standard score of the group on average was 67 (*SD* = 10.08, range=43-88), which corresponds to mild intellectual disability according to the classification of the DSM-IV-TR Manual (American Psychiatric Association, 2000) . As scores under 40 represent a severe intellectual disability, participants scoring below 40 were consequently removed from the sample.

(b) Reading comprehension level equivalent or superior to the third grade of elementary school (aged eight to nine). This inclusion criteria was selected to ensure that a minimum level of reading skill was acquired. This skill was measured through a standardized test called “Reading Comprehension Test” (ECL, De la Cruz, 1999) for Primary School. The test is composed of 2 levels, Level 1 for 2nd and 3rd graders and Level 2 for 3rd to 6th graders. For students with reading difficulties, the manual of the test recommends the use of the Level 1, so this was the version used in our study. The ECL1consistedof three short texts (approx. average length of 100 words) extracted from real school and literary Spanish books, accompanied by 17 literal and inferential questions. The application time

of the test is 30 minutes. The Cronbach's alpha of the ECL1 is 0.5219. The average direct score obtained was 8.96 (*SD* = 2.32.), which corresponds to the 50th percentile for the 3rd grade level. This test was administered by the three first authors of the present study, qualified by a degree in linguistic (the first one) and psychology (the other two).

Following the performance in these tests, one male participant was withdrawn from the sample for the score in the ECL1 was too low (direct score = 3 out of 17). The average age of the final sample was then 19.82 (*SD* = 2.16, *min* = 16, *max* = 24).

There was no main effect of gender for the Anaphor test scores (accuracy), t (26) = .415; p = .84), so this factor will not be considered in further analyses.

Control group. Twenty-seven (4 males, 23 females) Speech Therapy students aged between 20 and 25 (*M* = 21.6, *SD* = 1.31) from the University of Valencia voluntarily participated in this experiment as the control group. Reading level and IQ were assumed to be regular since they were undergraduate students, so baseline measures were not applied in this group. The 27 undergraduate students received additional course credit for their participation. In this group, the main effect of gender was neither significant for the Anaphor test scores (accuracy), *t* (25) = .834; *p* = .37, so this factor will not be considered further.

In both groups, students and/or students' parents signed an informed consent before participating in the study.

Materials. ***Anaphor decision task.*** For the anaphor decision task, the original stimuli designed by Gutiérrez-Martínez, García-Madruga, Carriedo, Vila, and Luzón (2005) , also employed in Elosúa et al.'s (2009) study were used here.

These conformed a pool of 84 sentences containing an object pronoun (42 morphosyntactically and 42 semantically related to the correct answer), followed by two answer choices presented in line, one to the right and one to the left of the screen. Elosúa et al. et al. followed some guidelines to create the sentences: (a) each sentence was 10-to-12 words long; (b) the answer choices were two- or three-syllables high frequency words; (c) only "la" (fem.) and "lo" (masc.) were used as targets, but not "le" (gender unmarked indirect object) to avoid a possible conflict due to a well-known misuse of that pronoun in Spanish called "leísmo"; (d) the order of the answer choices was counterbalanced, so that the correct answer and the distractor occur equally on the left and the right positions.

The sentence "Carmen **la** (f) encontró en el aula al lado del pupitre" (*Carmen found **her** in the classroom, next to the desk)* is given as a real example. In the morphosyntactic condition, participants had to choose between "carpeta" (f) (folder), and "cuaderno" (m) (notebook). Both choices match the pronoun semantically, but only the correct answer, *carpeta*, also does morphosyntactically. In the semantic condition, the choices would be "carpeta" (f) (folder), and "profesora" (f) (teacher). This time the concordance is only established through semantics, as in both choices the gender matches the feminine pronoun but only *folder* makes sense in that context. This test was administered by the three first authors of this research with degrees in linguistic (the first one) and psychology (the other two).

Procedure. Two collective sessions were necessary for the experimental group to perform the tasks: (1) a first session to accomplish the control tests by the experimental group, and (2) a second session, common to both groups, to perform the experimental task itself.

A within-subject design was created to present each participant with a total of 42 non-repeated stimuli, 21 of each kind, in a completely randomized order, by means of a Visual Basic solution created by the experimenters. For methodological reasons we had to change the original between-groups design employed by Elosúa et al., as our sample size was quite smaller than theirs.

In the second session participants were asked to sit on a computer individually and pay attention to the screen. They completed six practice trials followed by the 42 experimental trials. First, an instructions screen appeared to explain the procedure. By pressing the *next* button the first sentence would appear containing an object pronoun in bold. This sentence they should read carefully and, when ready press the *choices* button. The screen then changed to present the two alternatives to be evaluated: two nouns of different gender in the case of morphosyntactically disambiguated pronouns, and two nouns of the same gender for semantic anaphors. The correct answer was the more likely noun to be the antecedent of the pronoun in bold. By clicking in what the participant thought it was the correct choice and then the button *next*, the next stimulus appeared. Re-reading of the sentences was not allowed, but, as explained above, six practice items preceded the actual experimental stimuli in order to guarantee the comprehension of the procedure.

Results

The software employed to administer the Anaphor test also registered participants' choices (accuracy) and times (ms), these defined as the time employed in reading plus answering. Only times of the correct responses are

analysed. Mean values and standard deviations for both variables are shown in Table 1.

A first analysis to check the normality and homocedasticity principles revealed a no-normal distribution of the data, so non-parametric analysis were finally performed. For the within-group analysis the Wilcoxon signed-rank test was performed, whereas Mann-Whitney U served for the between-group comparisons.

Additionally, Pearson correlations were performed to explore a possible lineal dependence among the experimental task's results and the control tests, i.e. IQ and reading comprehension. We used an alpha level of .05 for all statistical tests

Table 1.

Experiment 1. Average response times (ms) and accuracy (*ms*), (*SD*).

	Response times		Accuracy	
	Morpho.	Sem.	Morpho.	Sem.
Experimental group	18371 (9034)	15064 (6833)	66.96 (16.92)	6621 (14.11)
Chrono. age-group	7236 (1592)	6788 (1560)	95.11 (5.5)	95.04 (5.6)

Accuracy. First, we tested the hypothesis that accuracy in the Anaphor test in the experimental group was not as good as in the control group (H1). Between-group comparisons yielded some significant results stating that the control group performed significantly better in both the morphological condition ($U = 47$, $p < .001$) and the semantic condition ($U = 37.5$, $p < .001$).

Following our H2, significant differences between conditions were expected in the group of students with ID but not in the control group. Within-group analyses

revealed a very short difference between the average percentage of correct responses in the morphologically and the semantically disambiguated pronouns, in both the experimental (morphological, *M* = 66.96, *SD*= 16.92; semantic, *M*= 66.21, *SD* = 14.11) and the control group (morphological, *M* = 95.11, *SD* = 5.49; semantic, *M* = 95.04, *SD* = 5.59). This very short difference resulted non-significant (See Figure 1).

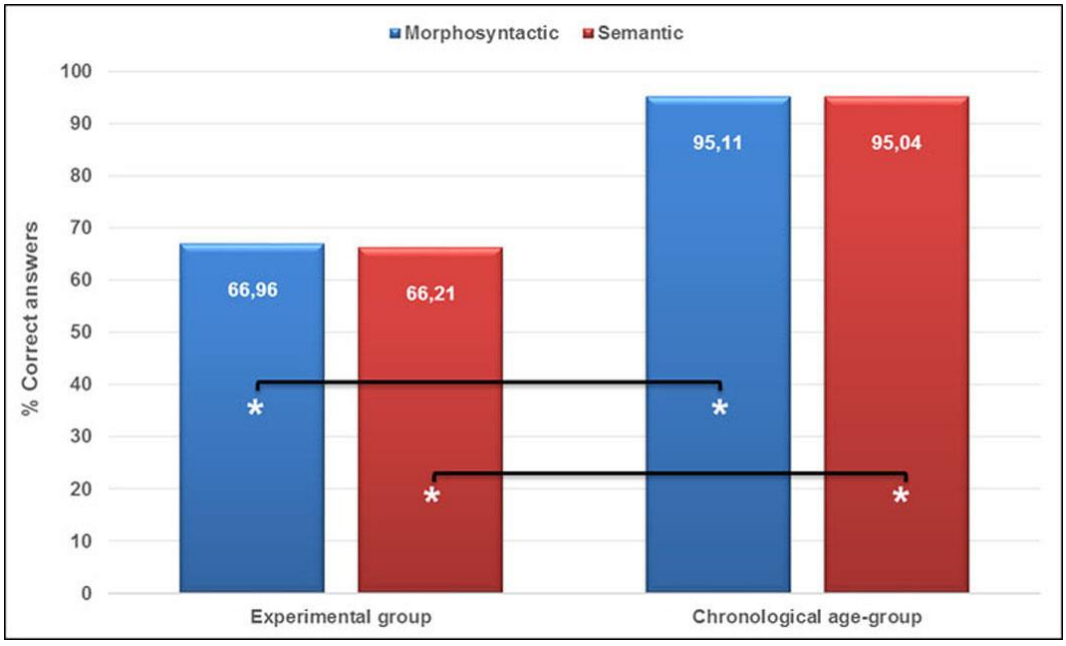

Fig. 1. Experiment 1. Percentage of correct responses in the pronoun resolution task, by group and condition (significant diferences are marked by an asterisk).

Response times. Though both groups followed the same trend, employing longer times in the morphological than in the semantically assigned antecedents, between-group comparisons were yet significant. Thus, in line with H1b, the students with ID employed significantly more time in the morphosyntactic condition

than their control peers (U = 40, p < .001), as well as in the semantic condition (U = 28, p < .001).

On the contrary to what happened with the percentage of correct responses, some significant results were found in the within-group analyses for the average time employed in the task. As we had hypothesised (H2), the experimental group dwelled longer when the disambiguation was morphosyntactic (M = 18370, SD = 9033) than when it was semantic (M = 15064, SD = 6833), and this difference resulted significant (Z = -4.21, p < .001). (See Figure 2).However, contrarily to our prediction results from the control group followed the same trend, with longer times in the morphosyntactic (M = 7235, SD = 1592) than in the semantic (M = 6788, SD = 1560) condition. Once again, this difference was significant (Z = -1.99, p < .05).

An additional analysis was performed to check whether the interference between morphosyntax and semantics was greater in the experimental group than in the control group. To this end, response times in the semantic condition were first deducted from the times in the morphosyntactic condition for each participant. A Mann-Whitney test served for analysing this new variable. As expected, the difference in the experimental group was greater (M = 3304, SD = 3599) than in the control group (M = 446, SD = 1005), and this difference resulted significant (U = 168, p < .00).

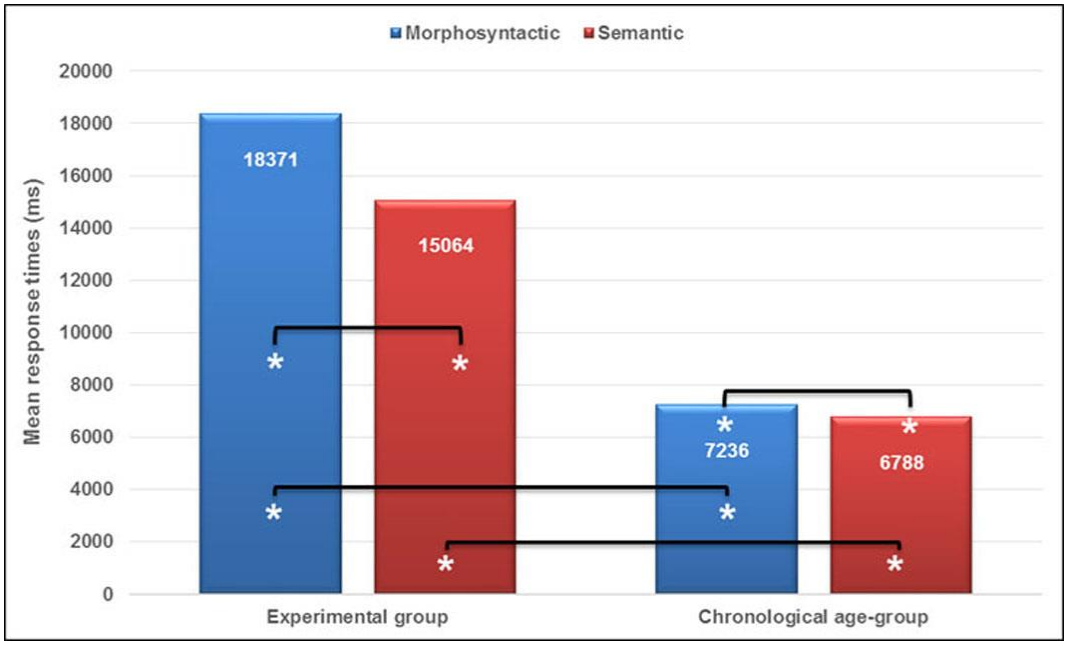

Fig. 2. Experiment 1. Avergage response times (ms), by group and condition (significant diferences are marked by an asterisk).

Discussion

In this experiment we tested the hypothesis that students with ID present a preference for following semantic cues when any conflictive information is introduced in a sentence. In this case, morphosyntax and semantics conflicted in a pronoun resolution task. As expected (H1a), the chronologically age-matched control group obtained higher scores than the ID students in both conditions. The times for resolving the anaphor were also hypothesised to be higher in the experimental group than in the control group (H1b), and this expectation was also accurate. Finally, a preference of semantics over morphosyntax in response times was found, but not only in the students with ID (H2), but also in the control group.

A possible explanation for the results of the control group, unexpected given the abovementioned literature, where an advantage of the morphological condition (with a pronoun gender cue) over semantic cues (Elosúa et al., 2009)

has been found, may be the design of the task. While in Elosúa et al's experiment the design was between-groups, in our sit was within-subject, which might have induced a different strategy to resolve the task, making readers to check both, semantic and morphological cues when both were available (morphosyntactic condition). This kind of strategy would have supposed a time penalty for both groups of readers, but especially for the experimental group, with less cognitive resources to invest in the task.

Experiment 1 did not provide the participants with an ecological environment in which they could read a complete text in the way they would do it in a normal daily situation. Presenting the complete text instead of using a sentence-by-sentence reading could also avoid memory overloads and prevent interferences between the observed variables and, especially, reading times. In 1992 O'Brien and Albrecht designed a methodology aimed to enhance the study of on-line comprehension processes, introducing thus the inconsistency detection paradigm that we use in the present experiment. This paradigm had similarities with previous studies (Baker & Anderson, 1982) , but it also supposes substantial changes such as the presentation of the whole text at a time. This study consisted of a set of three experiments, in which Psychology undergraduates read short passages of four to seven sentences. The first sentence introduced the main character and provided information about his or her spatial location (inside/outside). The critical sentence could be either the second or the fifth sentence and was led by a pronoun. The information provided in there could be consistent or inconsistent with the location of the protagonist specified in the first sentence. Apart from recording reading times, a comprehension question was presented after each passage to encourage participants to read for understanding.

A *distance* (close vs. distant) variable was introduced in experiment 1, while experiments 2 and 3 did not differ in the stimuli- similar to the close condition of the first experiment-, but in the instructions (adopt the perspective of the protagonist vs. read for comprehension). The results of experiment 1 demonstrated that readers did find the inconsistency and this fact was reflected by an increment of reading times in the inconsistent condition compared with the consistent for both distance conditions. The main effect of consistency was replicated in experiment 3, where specific instructions to adopt the perspective of the protagonist helped participants to detect the ambiguous information and caused them to dwell longer in the inconsistent passages. Oppositely, experiment 2 did not show any significant difference between the two conditions. Though the stimuli were the same as in experiment 3, the instructions were just targeted to reading comprehension what in light of the results, made a great difference.

Since their 1992's study, the design and procedure employed by O'Brien and Albrecht have been extensively used as a mean to obtain on-line and off-line processing measures at a time. This, the inconsistency detection paradigm, gives the researcher the possibility to track reaction (or reading) times, lexical decision times or naming times, among others, and study these data together with off-line comprehension measures such as correct responses to questionnaires. Moreover, researchers can track participants' monitoring skills during reading. It is precisely because of the possibility of obtaining such a complete outlook of comprehension, that the inconsistency detection paradigm has been frequently employed in the study of inference drawing and more relevant to our study, the comprehension of anaphors with ambiguous antecedents. The power of such a paradigm is considerably increased if it is used together with a technology like eye tracking, which on the one hand, renders a large amount of data that allows the researcher

to better profit from a single experiment; on the other, eye tracking provides information about assorted online cognitive processes that are otherwise hard to get to. For these two reasons it has been widely exploited (Just & Carpenter, 1980;Daneman & Carpenter, 1983; Traxler & Pickering, 1996; Rayner, 1997; J. Traxler, Bybee, & Pickering, 1997; Pollatsek, Fisher, & Rayner, 1998; Pearlmutter, Garnsey, & Bock, 1999; Traxler, Foss, Seely, Kaup, & Morris, 2000; Van Gompel & Pickering, 2001; Meseguer, Carreiras, & Clifton, 2002; Reichle, Rayner, & Pollatsek, 2003; Hyönä, Lorch, & Rinck, 2003; McDonald & Shillcock, 2003; Hirotani, Frazier, & Rayner, 2006; Pollatsek, Reichle, & Rayner, 2006; Rayner et al., 2006; Dubey, Keller, & Sturt, 2007; Staub & Rayner, 2012; Patterson, 2013) .

A study that profits the inconsistency detection paradigm and eye tracking is that of Rinck et al. (2003) which, by the way, is the one this experiment is inspired in. Their study consisted in two experiments in which they explore the processing of texts with temporal information when this is consistent (or inconsistent) with the situation model built up to then. Participants were undergraduates who performed a reading task while eye movements (first past fixations, second-pass fixations and regressions) were recorded. The texts to be read consisted of seven sentences; the second sentence introduced the temporal information, while the sixth presented an action either consistent or inconsistent with the information contained in the second sentence. In the first experiment the researchers set the focus on the time course of the processing of consistent vs inconsistent temporal information; in the second, they thoroughly explore how readers behave when those inconsistences are found and, hopefully, they try to recover from them. Additionally, participants in this second experiment had to detect and report the ambiguous information right after reading each text. As expected, readers from the first experiment fixated back the sentence containing the possible

disambiguating information more frequently than other areas, yielding more regressions and longer second-pass fixations in the inconsistent condition than in the consistent one. Interestingly, the same results were obtained in the second experiment, but only in those participants who were able to consciously report the inconsistence.

Besides Rinck et al.'s study (2003), another representative example of the use of the inconsistency paradigm together with eye tracking to explore anaphora comprehension is the work of Rayner et al. (2006). In it they presented two experiments, but the one that concerns us here is the second one. This experiment consisted on texts a bit longer (ca. 11 lines) than the ones in the original study by O'Brien and Albrecht, for which consistency (consistent vs. inconsistent) and distance between the antecedent and the anaphor (*close*-ca.12 words- vs. *middle*-ca. 55 words- vs. *far*- ca. 125 words) were manipulated. Instead of analysing the traditional measures of reading times and accuracy, they recorded the eye movements of American adult skilled readers to better understand the moment-by-moment operation of text processing. While no difference between the consistency conditions was encountered in the middle and long distances, the expected behaviour reflecting integration problems (i.e. longer fixations on the anaphor and more regressions to the antecedent) did appear in the closest inconsistent condition (ca. 12 words distant). These results suggest that the effect of the inconsistency diminishes as distance increases, very likely because participants update the possible antecedents as reading goes on. Rayner et al.'s results are not surprising, since distance has been previously proven to negatively affect anaphora resolution: as distance between the anaphor and its antecedent increases, the probability of considering a noun phrase the correct antecedent decreases (Duffy & Rayner, 1990; Myers & O'Brien, 1998).

Experiment 2 was designed on the basis of the antecedents on the inconsistency detection paradigm. The rationale, method and results of this second Experiment are described in the next section.

Experiment 2

The findings from our previous experiment evidenced the problematic of pronoun resolution experienced by youngsters with ID when there are competing antecedents. Just as predicted and based on previous literature (Megherbi & Ehrlich, 2005; Oakhill & Yuill, 1986) , there was an increment in readers with ID's response times mainly when the competing antecedents had to be evaluated and selected following morphosyntactic cues. The subsequent experiment kept on investigating anaphora resolution, following the work of Rinck, Gámez, Díaz, and De Vega (2003), this time using short texts instead of the single isolated sentences of the anaphor test by Elosúa et al. (2009). We put into play a text reading task from which, on the one hand, data of on-line processing of ambiguous pronoun was collected by recording eye movements. On the other hand, off-line processing was also tested through a questionnaire presented after the reading in which readers with ID, a chronologically age-matched group (young adults without ID), and a reading age-matched group (typically developing children), had to answer to comprehension questions reflecting different reading processes and report any inconsistence or ambiguity detected. This questionnaire consisted in three objective questions targeted to both literal and inferential information, one more question about the subjective perception of easiness and finally, the detection of the inconsistency itself. This questionnaire gave us the opportunity to explore the preference of use of semantic versus morphological cues in each group of readers.

Hypotheses

H1. Following the antecedents reviewed (Fajardo et al., 2013; 2014), students with ID are expected to experience problems in the inconsistent condition with the two inferential but not with the literal question. The ability to detect the inconsistency should be low, as it depends to a great extent on their metacognitive skills and these have been proven to be irregular in individuals with ID (Doğanay & Özmen, 2014). For the same reason, no difference between conditions is expected for the times and number of fixations after encountering the conflictive pronoun.

H2. The chronologically age-matched group would also perform equally in both conditions of the literal question since it is not affect by the anaphor inconsistency, but not in the inferential ones where lower accuracy scores are expected in the inconsistent than in the consistent condition. Overall scores are expected to be higher than in the experimental or in the children groups. For metacognitive skills are supposed to be normally developed, detection will also be better in this group than in their peers', and they will also dwell longer and with more fixations after reading the conflictive information than the other groups.

H3. At last, the reading age-matched group should obtain similar results to the ID students: lower accuracy in the inferential questions than in the literal question, in overall under the chronologically age-matched group scores. The inconsistency will not affect them greatly because they are supposed to be still developing their metacognitive and literacy skills, so times and number of fixations after the reading of the incongruent information should be similar in both experimental conditions.

Method

Participants.

An experimental group consisting of young students with ID and two control groups of students without ID participated in this experiment: (1) a chronological age-matched control group recruited at a mainstream primary school in Valencia (Spain) and (2) a reading level-matched control group. The experimental group was built up by some of the participants from the experimental group of the Experiment 1 plus some new candidates also recruited from the School of vocational Training *Camí Obert (*Valencian Community, Spain).

Experimental group. The inclusion criteria for participants with ID were similar as the ones in the previous experiment: (a) borderline to moderate intellectual disabilities and (b) reading comprehension level equivalent or superior to the third grade of elementary school. This time, twenty students with borderline to moderate intellectual disability were initially recruited from *Camí Obert*. Once again, the Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1997) was employed for measuring IQ, but reading comprehension level was assessed by means of the text *Penguins (*Cronbach's α= .653), which is one of the texts contained in the standardized TEC reading test (Vidal-Abarca et al., 2008). The original probe consists of two texts accompanied by 10 literal and inferential questions each, but we decided to use only one of the two texts because using the complete version could be exhausting for the students with ID. Following the K-Bit test battery scores one participant with a normal IQ was discarded, as well as 2 students with severe intellectual disability. The low comprehension rates in the reading task caused 4 more students to be pulled out from the sample (those with scores below the level of the 3rd graders). Additionally, recording errors were found in one participant's eye tracking data, who had to be withdrawn. Thus, the final

experimental group consisted of 12 students (4 male, 8 female) aged between 16 and 23 years (*M* = 19,6, *SD* = 2,54). The average IQ was 59,25 (*SD* = 13,3, range = 43-88), which corresponds with a mild intellectual disability according to the DSM-IV-TR Manual (American Psychiatric Association, 2000). Regarding the results of the TEC reading test, the group obtained a mean score of 4,25 (*SD* = 1,5, range= 3-7), out of 10 questions, corresponding to a percentile 50 in the 5th graders scale.

Chronologically age-matched control group. Nineteen students from different disciplines of the University of Valencia, in Spain voluntarily participated in the study as the chronological age-matched control group. Due to eye tracking errors three participants had to be withdrawn, so 16 students remained in the group (5 male, 11 female), aged from 18 to 20 years (*M* = 19,1, *SD* = 0,6). For they were all undergraduates, no additional tests were administered apart from the proper experimental task.

Reading level-matched control group. The reading age-matched control group first contained 25 5th graders from a regular Spanish primary school. Three of them were withdrawn for lacking some of the control tests, 2 because they were not native speakers of Spanish, which could interfere in the results, one because of the low scoring in the K-Bit, and 1 student more because of low comprehension rates in the TEC (Vidal-Abarca et al., 2008) reading task (only Text *Penguins* like in the experimental group was applied). This test was administered by the three first authors of this research qualified with degrees in linguistic (the first one) and psychology (the other two). Because of recording errors 6 participants more had to be withdrawn. Thus, a final control group of 12 children (4 male, 8 female) between 11 and 12 years old (*M* = 11,6, *SD* = 0,5) was used for the analyses. The IQ was measured again with the K-Bit intelligence test battery, obtaining an

average of 105.6 (*SD* = 15,1, range = 80-127). The mean direct score in the TEC reading test for this control group was 5,7 (*SD* = 1,8, range = 4-9), out of 10 questions, which corresponds to a percentile 60 in the 5th graders scale. A Student T-test confirmed that the difference between the percentiles obtained in the TEC in the experimental and the children control group was not significant (*t* (22) = -2,1, *p* = .06).

The main effect of gender for the accuracy scores of the inconsistency detection task was not significant for any of the groups: experimental, *t* (14) = .636, *p* = .54; chronological, *t* (16) = .70, *p* = .50; reading level, *t* (17) = .15, *p* = .88. That means, this factor will not be taken into account further on.

Participants or their legal tutors signed an informed consent before starting the study.

Materials. ***Inconsistency detection task*.** Four texts of six sentences were designed taking Rinck et al.'s (2003) materials as model. Sentences 1, 3, 4 and 6 were neutral, that is they contained no relevant information for the task though two proper names of different gender –masculine and feminine- were presented in the first one. Sentences 2 and 5 were the targets. Sentence 2 stated an action involving the two actants from sentence 1, always being one of them the subject and the other the indirect object. Sentence 5 was introduced by a personal pronoun in the syntactic role of the subject, whose antecedent could be either of the two proper names from sentence 2. (See Appendix A for a text sample).

The study followed a quasi-experimental design with *anaphoric consistence* (consistent vs. inconsistent) as the only independent variable and *pronoun gender* (masculine vs. feminine) as a non-experimental variable only used to control for

gender bias. Thus, four different versions were created from each text: consistent (half masculine, half feminine) and inconsistent (half masculine, half feminine). The (in) consistency was formed by matching the pronoun and the information contained in sentence 5 with the subject of the second sentence. In the *consistent* condition the pronoun-subject matching was semantically and morphosyntactically established (e.g. Pedro-he).On the contrary, the *inconsistent* condition happened when that matching was both semantically and morphosyntactically incongruent (e.g. Elena-he). If the pronoun (for instance, "he") did not match the subject of the second sentence ("Elena"), it was possible to attach it to the indirect object ("Pedro"), though the semantics of the context never allowed such a tie.

In addition, text length was controlled between conditions (range: 311-357 words).

The assignment of target texts to experimental conditions was counterbalanced across participants through four lists so that one participant read each experimental story just once and in only one experimental condition. Across participants, each text occurred equally often in each condition. The order of presentation of the four stories was randomized across participants by the SMI Vision design and presentation software *Experiment Center*. After each stimulus, the corresponding set of five comprehension questions appeared on screen one-at-a-time to test for global comprehension and conscious pronominal ambiguity detection.

Comprehension questions. As the goal of the study was to test and examine not only the online but also the offline processing of pronominal anaphora, especially when incongruences in the assignment of an antecedent occur, a set of five questions were designed to be presented after each text: three objective (*non-integrative, antecedent comprehension question,* and *anaphor integration*

question) and two subjective questions ("is there anything strange in the text?", "is there anything difficult in the text?") were presented on screen to be answered orally after each text (see Appendix B for a complete questionnaire sample). The subjective questions were recoded into two new variables: *explicit error detection* and *nature of the information provided*.

Literal question. The first inquiry referred to the first sentence and thus, it did not imply either the resolution of the anaphora or any interclausal integration (e.g. Where do Pedro and Elena live?).

Antecedent comprehension question. The second question asked about the information contained in the second sentence (for example, "Who phoned to meet on the beach?"). This time a total comprehension of the action (what happened?) as well as the role of the actants in it (who did what to whom?) was necessary.

Anaphor integration question. To answer this third question participants should resolve the anaphor, attaching the pronoun in sentence 5 to the most plausible antecedent in sentence 2 (e.g. "who always calls to meet?") From the point of view of the scoring, the consistent condition was clear and simple. The inconsistent condition, on the other hand, was arbitrarily scored on the basis of a morphosyntactic matching in which 1 (one) was assigned to the answer matching the gender of the pronoun and 0 (zero) to the semantic alternative (i.e. Pedro-he = 1; Pedro-she = 0). In other words, using a semantic attachment would sum 0 points, while the morphosyntactic strategy would add 1 point.

Explicit error detection. Participants answered here the question "do you see something strange/odd in text? This was a simple boolean variable with 0 (zero) as incorrect and 1 (one) as correct response.

Nature of the information provided. According to the answers to the subjective questions, the kind of information provided by participants was recoded into a three-level categorical variable where 0 (zero) meant no information provided, 1 (one) represented any kind of *morphosyntactic* statement (e.g. "*Pedro* cannot be *her*, that is wrong"), and 2 (two) was assigned to *semantic* information (e.g. "if Elena knows so much about files, why is Pedro the one resolving the doubt?").

Eye movements. To compute and analyse the dependent variables in eye tracking experiments it is necessary to select the target information in order to group the eye gaze data and thus ease the task of interpreting the results. Each group of eye gaze data is commonly called area of interest, henceforth AoI. In this case, six AoIs were first defined corresponding to the six sentences in the text. After several attempts to obtain any meaningful and explainable results, finally thicker measures had to be employed with no regard to any AoI distinction, mainly because of the irregularities in the eye movements of the participants with intellectual disabilities. Thus, the overall data for each text was finally computed and analysed.

Using eye tracking to study on-line comprehension processes allows the recording of a large amount of data. For this reason, grouping this data into *eye tracking measures* is essential (Rayner, 1997). For the present experiment several measures were computed for both early and late comprehension processes. As early measure, *first pass duration* reflects both the initial word recognition and the lexical processing. It is defined as the sum of the durations of all fixations in a region from first entering the region until leaving it either to the left or the right, given that the region was fixated at least once. As late measures, associated to integrative processes during reading comprehension, the number and duration of

the fixations made after leaving the critical area of interest containing the pronoun (sentence 5) were accounted, henceforth *fixations after offset S5* and *duration after offset S5*. This test was administered by the first author of the present research, holding a degree in linguistics and expertise in eye-movements measurement.

Procedure. A total of two sessions were needed to accomplish the tasks. Participants from the experimental group and the reading level-matched control group needed a first collective session in order to complete the K-Bit battery test and the TEC reading test.

In a second individual session participants fulfilled the Inconsistency Detection task. In this second session, the SMI Experiment Center software administered the task while a SMI RED250 stand-alone Eye Tracker was employed to record participants' eye movements throughout the task. Responses to comprehension questions were registered by the experimenter.

After giving the participant instructions to seat comfortably and to stay as quiet as possible during the task, the eye tracker was adjusted for optimal tracking. Participants were calibrated by a standard 9-point grid followed by another verification 9-point grid. They were instructed to read a set of short texts carefully and answer the questions verbally. The task was self-paced, so they could decide when to skip the screen and move on to the next question or text. Eye movements were recorded only during the reading of the text, not the questions.

Before each story, a fixation cross appeared centered on the screen surrounded by an invisible trigger-zone, so that after a 1000 ms gaze inside the boundaries the software automatically triggered the next stimulus. The next screen after reading the text appeared by pressing any key. At this point the student

should read the question on screen and answer verbally. A pause could be done whenever the participant asked for it. They completed 2 practice trials followed by the experimental items. The experimental trials did not start until the experimenter verified that the procedure had been understood.

Results

For the analysis of questions no participant was excluded because of recording errors, as these only appeared in eye tracking. Instead, the reading skills and IQ did serve as filters. The sample here was greater then, constituted by 16 participants in the experimental group, 18 in the chronologically age-matched readers' and 19 in the reading level-matched group. Non-parametric analyses, Mann-Whitney U for between-group comparisons and Wilcoxon matched pairs for within-group analyses, were performed to analyse the objective questions and eye tracking data. The variable *explicit error detection* was analysed by means of a Student-T, and finally, a Fisher's exact test served to explore the variable *nature of the information provided*. Data from the questionnaire is shown in Figure 3 and Figure 4.

Regarding eye movements, a filtering of outliers was performed prior to the analyses, considering outliers the times two standard deviations over or under the mean of each participant and condition. These values were replaced by the average time of the participant per fixation and they constituted a 3.5% of the values.

An alpha level of .05 was employed for all statistical tests.

Experimental group. No difference between conditions was encountered in the *literal question* or the *antecedent comprehension question*. Regarding the

anaphor integration question, the ID group obtained higher scores in the consistent (*M* = 84) than in the inconsistent condition (*M* = 34), *Z* = -3.56, *p* < .001. The *nature of the information provided* did not yield any significant difference, though "nothing" was reported more times than "morphosyntactic" or "semantic" (nothing, *N* = 21 vs. morphosyntactic, *N* = 5, semantic *N* = 6). Regarding eye movements, nor *first pass times*, *number of fixations after the offset of sentence 5*, or *duration of the fixations after the offset of sentence 5* reported any difference between the consistent and the inconsistent condition.

Chronologically age-matched control group. Again, results from the *literal question* reported no significance between conditions, and these were also close to a ceiling effect. This time, results from the *antecedent comprehension question* did not follow the same trend as in the experimental group, with significantly higher scores in the consistent (*M* = 97) than in the inconsistent condition (*M* = 50), *Z* = -3.69, *p* < .001. The *anaphor integration question*, on the other hand, did so, showing higher scores in the consistent condition (*M* = 100) than in the inconsistent (*M* = 47), and resulting this difference significant, *Z* = -3.95, *p* < .001. Fischer's exact test revealed a significant effect of *kind of information* in the *nature of the information provided* variable, *p* < .001. Besides, the trend was different as the other groups' (nothing, *N* = 8, morphosyntactic, *N* = 25, semantic *N* = 3). In line with the experimental group, eye tracking measures did not yield any significant result.

Reading age-matched control group. Results from the *literal question* were unexpected here, for children performed significantly better in the consistent (*M* =100%) than in the inconsistent condition (*M* = 87%), *Z* = -2.24, *p* < .03. Following the same trend as the chronologically age-matched group, the *antecedent comprehension question* in its consistent condition yielded a higher

percentage of correct answers (*M* = 95) than the inconsistent condition (*M* = 76), and this difference was significant, *Z* = -2.65, *p* < .01. The same happened in the *anaphor integration question*, with a greater preference for semantic cues in the consistent condition (*M* = 95) compared to the inconsistent (*M* = 53), and resulting the difference significant, *Z* = -2.24, *p* < .001. When it comes to *nature of the information provided*, results in the children's group were similar to the students with ID (nothing, *N* = 21, morphosyntactic, *N* = 10, semantic *N* = 7), showing no significant effect of *kind of information* at all. Once again, eye tracking data reported no difference between conditions.

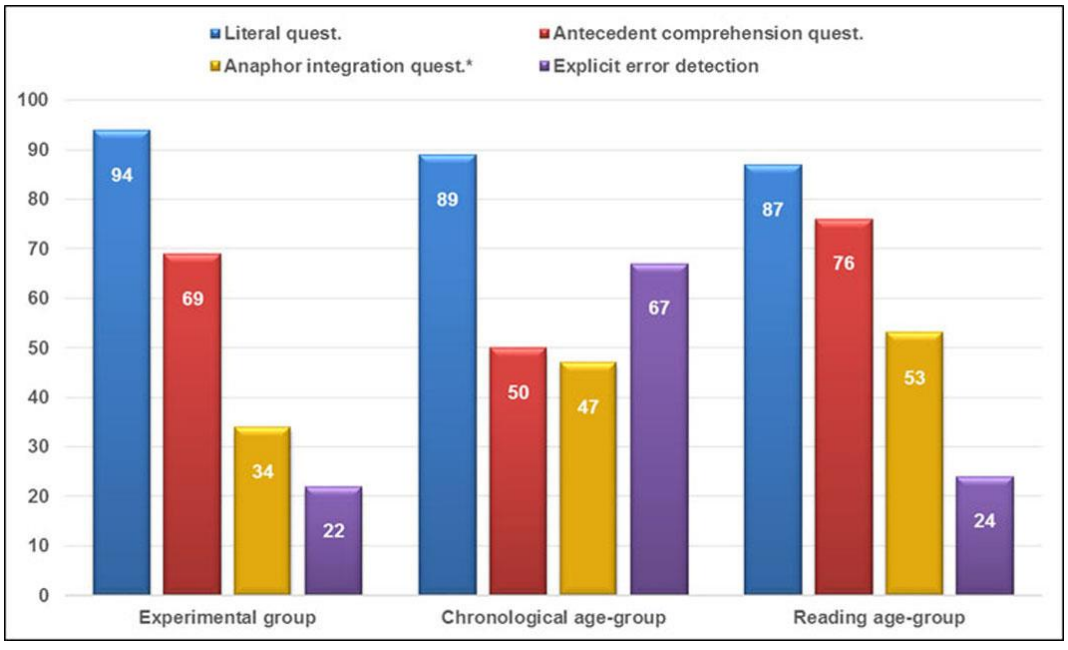

Fig. 3. Experiment 2. Average scores in the questionnaire, inconsistent condition by group (*average morphosyntactic matching).

Between-group. As we had hypothesised, no difference among groups was encountered in the *literal question* (see Figure 3 for a comparative view of the

scores in all questions for the inconsistent condition). However, some differences appeared in the performance for the *antecedent comprehension question*. Children performed better than the chronologically age-matched group in this question, $U = 91$, $p < .02$, but these results did not replicate when comparing to the ID students. The *anaphor integration question*, by its part, did not report any significant difference among groups, as all of them performed similarly: over 80% semantics in the consistent condition (morphosyntactic cues followed 20% of the times), and 60% in the inconsistent (morphosyntactic preference 40% of the times). Regarding the *explicit error detection*, just as expected, the chronologically age-matched group performed significantly better than the experimental group, t (32) = -3.30, $p < .003$, and the children, t (35) = -3.24, $p < .004$, (see Figure 4). An interaction between *group* and *kind of information* was found, $p < .001$, very likely because of the different results of the chronologically age-matched group compared with the experimental group and the reading level-matched group, concretely in what to the morphosyntactic category refers.

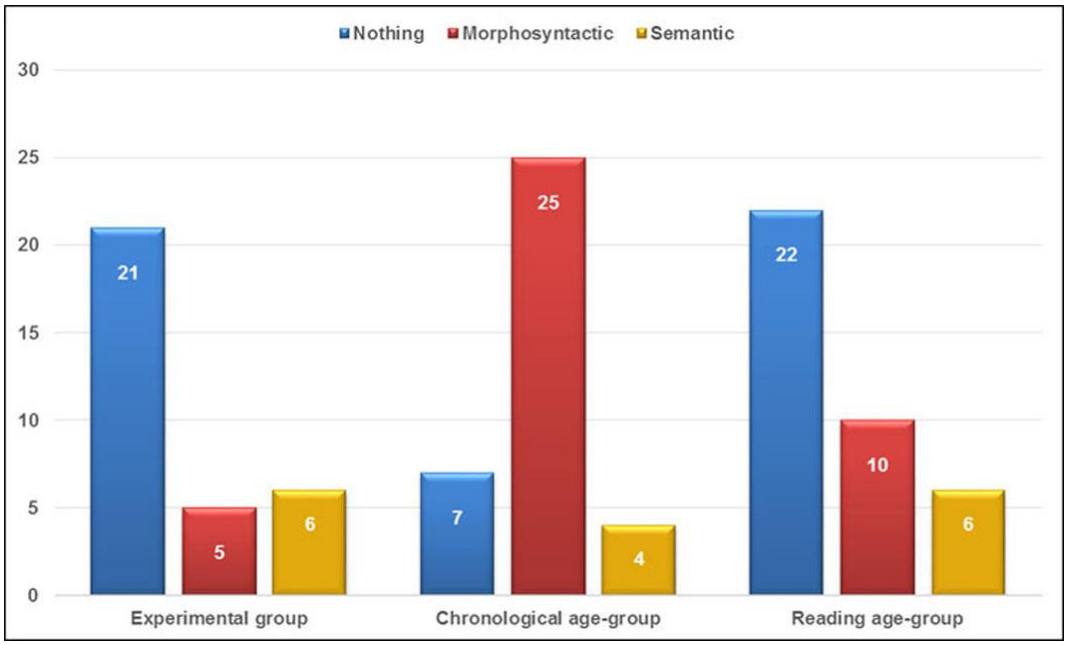

Fig. 4. Experiment 2. Kind of information reported (N) in the inconsistent condition by group.

The *duration of the first pass* was similar across groups, so analyses did not return any significance. The *number of fixations after the offset of sentence 5*, on the other hand, yielded some significant results. As we had hypothesised, normative adults made more fixations after the pronoun than the experimental group in both the consistent (chronologically age-matched, *M* = 33.97; experimental group, *M* = 11.33) and the inconsistent conditions (chronologically age-matched, *M* = 33.75; experimental group, *M* = 21.21). In both cases the difference was significant (consistent, *U* = 23.5, *p* = .001; inconsistent, *U* = 52, *p* = .04). Additionally, the children were encountered to refixate the text more than the ID students in the consistent condition (reading age-matched, *M* = 32.58; experimental group, *M* = 11.33), and this difference was significant, *U* = 26, *p* < .01. Regarding the *duration of the fixations after the offset of sentence 5*, some unexpected results were obtained, as the chronologically age-matched group dwelled significantly longer in re-reading the text when no inconsistence was introduced than the ID students (chronologically age-matched, *M* = 6095 ms; experimental group, *M* = 2545 ms), *U* = 33.5, *p* < .01. The same happened in the reading age-matched group in comparison with the experimental group (reading age-matched, *M* = 6882 ms; experimental group, *M* = 2545 ms), *U* = 32, *p* < .03.

Discussion

In this second experiment we re-tested the hypothesis of the semantic vs. morphosyntax advantage in pronoun resolution by readers with ID, this time presenting a pronominal anaphor in a text in an inconsistency detection task.

For the experimental group we expected low accuracy in the inconsistent condition of the inferential questions (H1) but this was only partially accomplished.

These results were obtained only in the *anaphor integration question*, showing a preference for the semantic pronoun-antecedent matching, but they performed equally well in both experimental conditions of the *antecedent comprehension question*. A possible explanation is a failure in their metacognitive skills, so that they do not notice the ambiguity answering this question in both conditions. Similarly, both eye movements and *explicit error detection* scores showed this same poor metacognitive performance. Regarding the kind of *information provided* about the nature of the ambiguity, the students with ID reported "nothing" in most of the cases, while "morphosyntactic" and "semantic" obtained rather similar results, showing thus no preference for one or another strategy.

The chronologically age-matched control group was expected to show clear differences between the consistent and the inconsistent conditions in the two inferential questions (H2), and so did they, with a better performance in both cases in the consistent condition. In other words, they were actually sensitive to the anaphoric inconsistency. Regarding the prediction about the preference of a semantic strategy, scores in the *anaphor integration question* were close to 50%, reflecting no preference of such a strategy over the morphosyntactic one. *Explicit error detection* was high (ca. 70% correct detections), but eye movements did not show any change in the reading strategy from the consistent to the inconsistent condition. About the *nature of the information provided*, the control adults reported the inconsistency far more frequently to be a "morphosyntactic" issue than "semantic" or "nothing". This result is consistent with the other variables, supporting the hypothesis that normative adults possess a better metacognitive ability.

Following our hypothesis H3, the children group should have had similar results to the ID students, reflecting low metacognitive skills in any case.

Surprisingly, they obtained higher scores in the consistent condition not only in the *antecedent comprehension question*, but also in the *literal question*. As children also obtained scores close to 50% in the *anaphor integration question*, we draw the same conclusion here as for the chronologically age-matched group: they show no preference for any of the two tested strategies. As expected, *explicit error detection* was very low and eye movements were similar in both experimental conditions.

Comparing the groups, the chronologically age-matched control group also demonstrated to owe better metacognitive skills. Even though no significant result was obtained in the between-group comparisons for the *literal question* or the *anaphor integration question*, they obtained lower scores in the *antecedent comprehension question*, very likely due to the impact of the ambiguity. Children seem to have realised about the inconsistency, but they probably evaluate it as not so important, so they could easily answer the question, which asked about information contained in sentence 2 (no integration of the anaphor was needed). The adults control group, on the contrary, did completely notice the ambiguity, so they made it extensive to the comprehension of the antecedent, experiencing problems to answer this question. Conscious detection was also better in this group. When it comes to eye movements, no group spent more time than the others in any condition. However, both children and normative adults made more refixations after reading the last sentence than the ID students, meaning that the default reading strategy of both control groups was qualitatively different than the ID students'.

General discussion

For this study we designed two experiments to explore anaphora resolution under conflictive circumstances. In Experiment 1 we adapted the anaphora test of Gutiérrez-Martínez et al. (2005) and Elosúa et al. (2009) to test how students with ID and a chronologically age-matched group of undergraduates performed a task involving anaphors, morphosyntax and semantics. Participants should decide, between two answer choices, which word a clitic pronoun contained in a given sentence referred to. Half of the times, the correct answer could be unravelled through a gender cue, while the other half it was the semantics of the context and the word itself what served to that end. Both groups performed equally well in the two experimental conditions. However, the times employed in resolving the morphosyntactic anaphor were significantly higher than the semantic condition, in both groups. As expected, the normative adults performed in overall better than the ID students.

Experiment 1 inspired us to continue our investigation in the line of the contrast morphosyntax vs. semantics. Hence, we designed Experiment 2 inspired in Rinck et al. (2003) and the inconsistency detection paradigm to assess both on-line and off-line processing of anaphoric pronouns and detection of binding inconsistences. Participants read four texts whose second sentence contained two proper names of the same gender (in subject and indirect object roles) while the fifth sentence was introduced by a pronoun bound to one of those names, half the time the subject, half the time the object (in this case, overall semantics resulted ambiguous). To avoid gender bias, as it has been proven to influence pronoun resolution (Arnold et al., 2000), half the names were masculine and half feminine. By recording eye movements we wanted to obtain on-line evidence about the recognition and integration processes of ambiguous and unambiguous pronouns.

An additional questionnaire would allow us to collect data about participants' awareness of this ambiguity, their objective disambiguating accuracy and their literal text comprehension.

The results obtained in Experiment 1 were similar to Elosúa et al.'s in accuracy, as no difference between the experimental conditions was found in both groups. Though overall mean accuracy was lower in the ID students than in the control group, no difference at all was found. These results are consistent with the fact that the stimuli were originally conceived to serve in an operative memory test, so they were created intentionally easy for young adults.

Finding an explanation to processing times seems trickier. While Elosúa et al. obtained evidence supporting the superiority of morphosyntactic analysis, reflected by shorter processing times, our results go in the opposite direction. The reason why our students rely on semantics instead of morphosyntax may lay on the design itself. Elosúa et al. manipulated the experimental conditions between-groups, meaning that every participant was exposed to only one condition, either semantic or morphosyntactic. Our design, on the other hand, was a within-group, so that every participant read the same number of items of each condition and the items of each condition were mixed in such a way that within the same trial block participants resolved both morphosyntactic and semantic trials presented in a random order. This fact could be having an impact on the results in what to strategies refers: as they realize that gender cues are only sometimes present, they decide to employ a semantic strategy by default. When the anaphor is to be semantically disambiguated, response times are shorter than Elosúa et al.'s control group. On the contrary, when they are supposed to employ morphosyntactic cues but semantics are used instead, a penalty in time appears as switching from semantics to morphosyntax supposes a delay.

As the participants in Elosúa et al.'s study were young adults with no explicit literacy problem, it is understandable that the results obtained by our ID students remain unexplained from the same perspective. However, the results obtained in our Experiment 1 for this group are similar to Megherbi and Ehrlich's (2005). They observed that only skilled readers were able to take advantage of morphosyntax when gender cues are available besides semantics. Low-skilled children, on the other hand, preferred to resolve the anaphor following the semantics of the verb instead of the morphosyntactic cues. These results could explain why the children and ID students in our experiment seem to prefer the semantic disambiguation as default strategy, even though taking advantage of a surface element such as gender signs could aid a faster resolution of the anaphor (Cacciari et al., 1997).

For Experiment 2 we argued that students with ID would experience problems with the two inferential questions (H1), but the results did not completely accomplish this expectation. In fact, only the question asking directly about the integration of the anaphor resulted more intricate in its inconsistent condition. Consequently with our predictions, though the inconsistency did cause some problems in the resolution of the anaphor (34% correct responses in the *anaphor integration question*), students with ID neither notice it consciously (22% correct detections) nor changed their reading strategy to re-read the incongruent parts of the text (no difference between conditions in eye movements). These results were in overall what we expected, as in the literature review we had found enough antecedents reporting the limited metacognitive skills of individuals with ID (e.g. Doğanay & Özmen, 2014). Besides, given the low reading comprehension level of the group, it was anyway reasonable to anticipate a behaviour proper from low-skilled readers in terms of a fairly detection and a poor reading strategy.

The chronologically age-matched control group encountered some problems in the two integrative questions, just as we hypothesised (H2). Given the nature of the *antecedent comprehension question* and the scoring criteria of the *anaphor integration question,* we can assert that these scores, in both cases close to 50%, reflect no preference of semantics or morphosyntax, for they seem to use both equally. Once again, this finding is consistent with Megherbi and Ehrlich's, who found skilled readers to take advantage of both cues in a pronoun resolution task. Precisely for this reason, it is a bit puzzling that participants from this group reported the inconsistency as a morphosyntactic far more frequently than a semantic issue. It may be that they consciously know the grammar rules of the language and after detecting that the problematic of the text lays on the pronoun (67% correct detections), they get to the conclusion that grammar should be responsible of such incongruence. A bit surprising is the lack of effect of the inconsistency in eye movements, because this group was expected to be affected by it but, on the contrary, they followed the same reading strategy in both kinds of texts.

Regarding the reading age-matched results, our hypothesis 3 was again quite accurate. Children showed some problems with the inferential questions when the anaphor was not consistent within the context. However, they also performed much better in this condition when a simple literal question was made. For any reason, they seem to encounter the text globally inconsistent when an inconsistence is introduced, even though this was supposed to affect only to the integration of the sentences 2 and 5 (*anaphor integration question*). Detection skills were very low (24% correct detections), similarly to the ID students', just as we had expected. Again, the low detection rates together with the absence of effect of the inconsistency in the eye tracking data reflect impoverished

metacognitive skills that prevent children from performing an adequate reading strategy.

When we compared the scores and behaviour of the three groups some outstanding results came to light. Very interestingly, a main effect of inconsistency was encountered in the *antecedent comprehension question* for both control groups. Even though they performed significantly better in the consistent condition, children's accuracy was higher than normative adults'. This apparently surprising result might not be so rare, given that the questionnaire was presented once the text had already been completely read and the chronologically age-matched group, who in the light of the results have greater inconsistency detection skills, might be more affected by the ambiguity than the children.

When it comes to the explicit error detection, the chronologically age-matched group did perform significantly better than the children and the ID students, what completely matches our expectations.

Taking all the results together, we could argue that children as well as normative adults are more likely to re-read the text than the students with ID as a default strategy, maybe to check whether the situation model they have constructed through the first reading is trustworthy (Baker & Anderson, 1982) . Nonetheless, when an inconsistency is introduced only the chronologically age-matched seem to detect it, making more refixations than the experimental group. This behaviour posits the use of different reading strategies in the typically developing individuals and youngsters with intellectual disabilities, although apparently the normative adults' is the most efficient strategy, for they are able to detect the conflictive information more frequently than the other participants.

Regarding the *nature of the information provided* and taking the results of Experiment 1 into account, a higher count of semantic explanations were to be expected in the ID students and possibly the children, but not in the chronologically age-matched control group, where no preference for a certain strategy was hypothesised. This expectation was not completely achieved. Instead, a substantial count of *morphosyntactic* explanations were given by the chronologically age-matched group in comparison with the *semantic* or the *nothing* categories, meaning that they regard the inconsistency as a question of morphosyntax and not so much of semantics. In the children and the ID students groups the trend was quite different: very few times they pointed to semantics or morphosyntax, but they reported much frequently *nothing*. We argue that the contradiction between these results and the questions' might be explained again by means of the impoverished metacognitive skills.

At last, we also hypothesised better detection results in the inconsistent condition in terms of a higher count of refixations and longer refixation times in the normative young adults in comparison with the children and ID students. What we observed was that, in fact, the adults without ID made more fixations in second and subsequent readings than their peers with ID. Also children did, but only in the consistent condition. On the one hand, this result is intended in the same line as the other variables, as a reflection of better metacognitive skills in non-ID adults than in ID students. On the other, it states that the strategies used by the three groups tested are qualitatively different: while children and adults tend to re-read the text, either including any ambiguity or not, the ID students do not check their own model of the situation. Taking this result isolated, it does not necessarily mean that they do not notice the inconsistency, as previous research has already given account of the "laziness" of low-skilled readers, who do not modify the gist of their

models even though they detect some inconsistent information (van Oostendorp, Otero, & Campanario, 2002; Otero & Campanario, 1990; Otero & Kintsch, 1992) . However, together with the questionnaire results, we should claim that the strategy of our ID students is poorly efficient, because they rarely review their model of the situation, and they hardly ever manage to detect the inconsistency.

Conclusions

The results obtained in these experiments give us a new insight into how the detection of inconsistencies works in the observed populations. We have seen that there exist little and isolated differences between Spanish 5th graders and normative young adults in what to reading strategies refers. Children are, however, not completely competent in detecting the inconsistency, as those strategies only resulted successful in the case of the young adults. Factually, youngsters with intellectual disabilities and children obtained similar off-line results despite the differences in the on-line processing. This is an interesting result from the point of view of a reading intervention: despite the isolated differences in the processing of ambiguous pronouns that can be found throughout this research, it seems that the level of awareness is rather similar in children and young adults with intellectual disabilities, so a proper intervention to train ID students in reading strategies might be worthy. In the framework of text design, the conclusions from this research come to say that including pronominal anaphora is not strictly a *bad practice* when text comprehension is to be enhanced, for the participants in our studies seem to have problems only when the antecedents of these anaphors are not clear or they are presented in an unnatural reading context (like the isolated sentences of Experiment 1) where there is few semantic information available to build the text representation (O'Brien & Albrecht, 1992) .

Despite reading skills are central in curricula of special education or inclusion situations, literacy acquisition is still limited for students with ID as recent studies have shown (Ratz, & Lenhard, 2013; van Wingerden, Segers, van Balkom, & Verhoeven, 2014). Thanks to the information provided by the present study we have learnt that training on the use of metacognitive skills should be incorporate to these curricula, instead of training the mere isolated decoding skills during reading

comprehension. Actually, teaching to use metacognitive strategies like the activation of prior knowledge have been successful to increase inference making during reading in students with ID (Morgan, Moni, & Jobling, 2004). However, it also seems that the processing of morphological cues does not occur automatically in people with ID, who prefer to use semantic cues even though this could eventually be reducing the available memory resources needed to put other metacognitive skills into play. This reduction of the available memory resources during reading could be especially relevant for students with ID due to their low short term memory capacity (Van der Molen, Henry & Van Luit, 2014). Therefore, an alternative or even parallel strategy to increase reading comprehension is suggested to improve morphosyntactical skills. This alternative to traditional training programmes on reading could have a direct beneficial effect on reading comprehension but could also liberate resources devoted to metacognition. This solution could be especially interesting for some aetiologies of ID like Down syndrome, which seems to present a specific delay on syntax acquisition (Fowler, 1990).

Acknowledgements

This research was funded by the AVANZA intellectual disability program (Project expedient: TSI-040200-2008- 0063) of the Spanish Minister of Industry, Tourism and Trade and Technosite (ONCE Foundation Group). We thank the participants from Camí Obert, Colegio La Fontaine and the University of Valencia for their kind and valuable collaboration in this study. Workers of the Camí Obert and APSA organizations, as well as the students' families have also played a fundamental role allowing and facilitating the access to the students and other resources needed.

References

American Psychiatric Association. (2000). *DSM-IV*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Association.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22* (4), 527–565.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76* (1), B13–B26.

Baker, L., & Anderson, R. I. (1982). Effects of Inconsistent Information on Text Processing: Evidence for Comprehension Monitoring. *Reading Research Quarterly*, *17* (2), 281.

Cacciari, C., Carreiras, M., & Barbolini, C. (1997). When Words Have Two Genders: Anaphor Resolution for Italian Functionally Ambiguous Words. *Journal of Memory and Language*, *37* (4), 517–532.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, *11*, 489–503.

Cain, K., Oakhill, J., Barnes, M. A., & Bryant, P. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29* (6), 850–9.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research*, *19* (4), 246–264.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9* (4), 561–584.

De la Cruz López, M. V. (1999). Evaluación de la comprensión lectora: ECL. Madrid: Tea.

Department of Education (2008). PCPI Regulating Ordinance (Publication No. 5790/23.06.2008). Valencian Community, Spain.

Doğanay Bilgi, A., & Özmen, E. R. (2014). The Impact of Modified Multi-component Cognitive Strategy Instruction in the Acquisition of Metacognitive Strategy Knowledge in the Text Comprehension Process of Students with Mental Retardation. *Educational Sciences: Theory & Practice*, *14* (2), 707–715.

Dubey, A., Keller, F., & Sturt, P. (2007). The Effect of Phonological Parallelism in Coordination Evidence from Eye-tracking. In *Proceedings of the 2nd European Cognitive Science Conference* (pp. 131–136).

Duffy, S. A., & Rayner, K. (1990). Eye Movements and Anaphor Resolution: Effects of Antecedent Typicality and Distance. *Language and Speech*, *33* (2), 103–119.

Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in

metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal*, *11*, 29–63.

Elosúa, M. R., Carriedo, N., & García-Madruga, J. A. (2009). Dos nuevas pruebas de Memoria Operativa de Anáforas. *Infancia Y Aprendizaje*, *32* (1), 97–118.

Fajardo, I., Ávila, V., Ferrer, A., Tavares, G., Gómez, M., & Hernández, A. (2014). Easy-to-read texts for students with intellectual disability: linguistic factors affecting comprehension. *Journal of Applied Research in Intellectual Disabilities : JARID*, *27* (3), 212–25.

Fajardo, I., Tavares, G., Ávila, V., & Ferrer, A. (2013). Towards text simplification for poor readers with intellectual disability: when do connectives enhance text cohesion? *Research in Developmental Disabilities*, *34* (4), 1267–79.

Fowler, A. E. (1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. *Children with Down syndrome: A developmental perspective*, 9, 302-328.

Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, *4* (4), 323–347.

Garvey, C., Caramazza, A., & Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, *3* (3), 227–243.

Gelormini-Lezama, C., & Almor, A. (2011). Repeated Names, Overt Pronouns, and Null Pronouns in Spanish. *Language and Cognitive Processes*, *26* (3), 437–454.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*, *17* (3), 311–347.

Gutiérrez-Martínez, F., García-Madruga, J.-A., Carriedo, N., Vila, J.-O., & Luzón, J.-M. (2005). Dos pruebas de amplitud de memoria operativa para el razonamiento (Two working memory measures for reasoning). *Cognitiva*, *17* (2), 183–207.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54* (3), 425–443.

Hyönä, J., Lorch Jr, R. F., & Rinck, M. (2003). Eye Movement Measures to Study Global Text Processing. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 313–334). Amsterdam.

Jakubowicz, C. (1991). L'acquisition des anaphores et des pronoms lexicaux en français. In J. Guéron & J. Y. Pollock (Eds.), *Grammaire générative et syntaxe comparée* (pp. 229–253). Paris: Editions du CNRS, Sciences du Langage.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16* (4), 260–4.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87* (4), 329–354.

Kaufman, A. S., & Kaufman, N. L. (1997). *Test breve de inteligencia de Kaufman (K-BIT).* Madrid: TEA.

Long, D. L., & De Ley, L. (2000). Implicit Causality and Discourse Focus: The Interaction of Text and Reader Characteristics in Pronoun Resolution. *Journal of Memory and Language*, *42* (4), 545–570.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14* (6), 648–52.

Megherbi, H., & Ehrlich, M.-F. (2005). Language Impairment in Less Skilled Comprehenders: The On-Line Processing of Anaphoric Pronouns in a Listening Situation. *Reading and Writing*, *18* (7-9), 715–753.

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30* (4), 551–561.

Morgan, M., Moni, K., & Jobling, A. (2004). What's it all about? Investigating reading comprehension strategies in young adults with Down syndrome. *Down Syndrome Research and Practice*, 9 (2), 37-44.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26* (2-3), 131–157.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18* (4), 777–84.

Oakhill, J., & Yuill, N. (1986). Pronoun Resolution in Skilled and Less-Skilled Comprehenders: Effects of Memory Load and Inferential Complexity. *Language and Speech*, *29* (1), 25–37.

Oostendorp, H. van, Otero, J., & Campanario, J. M. (2002). Conditions of updating during reading. In M. M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary Studies* (pp. 55–76). Amsterdam: John Benjamins Publishing Co.

Otero, J. C., & Campanario, J. M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching*, *27* (5), 447–460.

Otero, J., & Kintsch, W. (1992). Failures to Detect Contradictions in a Text: What Readers Believe Versus What They Read. *Psychological Science*, *3* (4), 229–235.

Patterson, C. (2013). Discourse Coherence in Pronoun Resolution. *Discours*, *12* (12).

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement Processes in Sentence Comprehension. *Journal of Memory and Language*, *41* (3), 427–456.

Perovic, A. (2006). Syntactic deficit in Down syndrome: More evidence for the modular organisation of language. *Lingua*, *116* (10), 1616–1630.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52* (1), 1–56.

Rademaker, A., & Haeusler, E. H. (2008). Toward Short and Structural ALC-Reasoning Explanations: A Sequent Calculus Approach. Proceedings of

Brazilian Symposium on Artificial Inteligence. Advances in Artificial Intelligence - SBIA 2008. 167-176. Springer: Berlin.

Ratz, C., & Lenhard, W. (2013). Reading skills among students with intellectual disabilities. *Research in developmental disabilities*, 34 (5), 1740-1748.

Rayner, K. (1997). Understanding Eye Movements in Reading. *Scientific Studies of Reading*, *1* (4), 317–339.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, *10* (3), 241–255.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105* (1), 125–57.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences*, *26* (4), 445–76; discussion 477–526.

Rinck, M., Gámez, E., Díaz, J. M., & De Vega, M. (2003). Processing of temporal information: evidence from eye movements. *Memory & Cognition*, *31* (1), 77–86.

Santiuste, V. (1997). Adquisición de anáforas reflexivas y de pronombres desde la perspectiva de la teoría de principios y parámetros. *Revista de Filología Románica*, *1* (14), 445–454.

Staub, A., & Rayner, K. (2012). *Eye movements and on-line comprehension processes*. *Language and Cognitive Processes* (Vol. 4, pp. SI21–SI49). Oxford University Press.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of Connectives on Language Comprehension: Eye tracking Evidence for Incremental Interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *50* (3), 481–497.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, *29* (6), 581–95.

Traxler, M. J., & P, M. J. (1996). Case-marking in the Parsing of Complement Sentences Evidence from Eye Movements. *The Quarterly Journal of Experimental Psychology*, *49A* (4), 991–1004.

Van der Molen, M. J., Henry, L. A., & Van Luit, J. E. H. (2014). Working memory development in children with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, 58 (7). 637-650.

Van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin & Review*, *8* (4), 851–857.

Van Wingerden, E., Segers, E., van Balkom, H., & Verhoeven, L. (2014). Cognitive and linguistic predictors of reading comprehension in children with intellectual disabilities. *Research in developmental disabilities*, 35 (11), 3139-3147.

Vidal-Abarca, E., Gilabert, R., Martínez, T., Sellés, P., Abad, N., & Ferrer, C. (2007). Test de Estrategias de Comprensión (TEC). Madrid: Instituto Calasanz de Ciencias de la Educación.

Yuill, N., & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, *79* (2), 173–186.

Appendix A

English translation of a sample text used in the experiment

1.	Pedro (m) and Elena (f) are two friends who live in Alicante
2a.	Consistent: *Elena phoned to Pedro to spend the day on the beach
2b.	Inconsistent: *Pedro phoned to Elena to spend the day on the beach
3.	Everything was almost ready, but the beach umbrella was really dirty and broken
4.	In a shop on the beach Elena and Pedro bought a new beach umbrella
5.	Target sentence: She always calls her friends to meet
6.	In summer they will go to the beach more frequently

Note—Each participant read either Sentence 2a or 2b. (m) = masculine; (f) = feminine.

Appendix B

English translation of the sample set of questions for the text in Appendix 1

Literal question	Antecedent comprehension question	Anaphor integration question	Subjective Question 1	Subjective Question 2
Where do Pedro and Elena live?	Who phoned to meet on the beach?	Who always calls to meet?	Is there anything strange in the text? (What?)	Is there anything difficult in the text? (What?)
Answer choices				
In Alicante	Elena	Elena	(OPEN)	(OPEN)
On the beach	Pedro	Pedro		
In Valencia	Nobody	Nobody		

6 General discussion and conclusions

6. General discussion and conclusions

This thesis has tried to delve into the use of adaptation and simplification of texts as a facilitator means for reading comprehension in young people with ID. As explained in the introductions of the three papers abridged here, the lack of consensus between the antecedents in the literature and the scarcity of studies focusing on the empirical validation of the adaptation guidelines proposed by international organizations such as IFLA, especially for the group of readers with ID, required such a work.

Thus, throughout three papers (5 studies) we have tried to answer two main questions:

1) Do simplified texts improve reading comprehension in young students with moderate-to-borderline ID? The selected measures of reading comprehension have been designed on the basis of the Construction-Integration Model by Kintsch (1998), distinguishing between literal (surface measure) and inferential comprehension (deep measure).

2) If they did boost comprehension, what psycholinguistic variables are involved in this improvement? Again with the aim of leading our predictions theoretically and operationalizing the measures used in the adaptation and simplification of texts, we distinguished between the adaptation at a surface level and the adaptation at a cohesion level (following the classification suggested by Crossley, Dufty, McCarthy and McNamara [2007]; Crossley, Greenfield and McNamara [2008] and Crossley y cols. [2007]). In rough outlines, the predictions pointed to a greater impact of the surface adaptations on literal comprehension, on the one hand, and a greater impact of the adaptations at the level of cohesion on inferential comprehension, on the other.

In this section we will summarise the results of the five studies while trying to answer these questions. Moreover, we will also describe their implications from an applied point of view.

6.1. The effect of the adaptation of texts on the reading comprehension of youngsters with ID

In this work we have witnessed two phenomena. Firstly, our sample of young readers with moderate-to-borderline ID had a reading comprehension level under their chronological ages', comparable to Spanish 2nd-3rd graders. Secondly, the adaptation of texts boosted reading comprehension indeed, especially at the literal level. In the first Study we observed how journalistic texts adapted following the guidelines of the IFLA made possible the literal and inferential comprehension (correct answers over 70%). Nonetheless, there was still a significant difference between them both, with lower accuracy rates in the latter. This result was replicated in Study 2.1, reflected by a significant higher literal comprehension without any statistical effect of the simplification though. No text adaptation was used in Studies 2.2, 3.1 and 3.2. Instead, we assessed the processing of assorted cohesion variables, examining the effect they have on the reading comprehension of our youngsters with ID. In the two next sections we will deal with this facet.

6.2. Linguistic measures that modulate/ predict the effect of the adaptation

6.2.1. Effectiveness of the adaptation at the surface level

In the three first studies we have observed that the adaptation guidelines based on surface measures (e.g. word frequency in Studies 1 and 2.1, word familiarity in Study 2.2, and Inflesz legibility index in Study 1), do not imply the main effects predicted by the theoretical models and the antecedents in the literature. For instance, previous researches had obtained positive outcomes regarding processing and comprehension by rising word frequency and/or word familiarity (Haberlandt & Graesser, 1985; Just & Carpenter, 1980; Rayner & Duffy, 1986; Crosson et al., 2008). However, our experience does not back up this improvement. Augmenting the lexical frequency showed no facilitative effects in Studies 1 and 2.1, albeit either a ceiling effect or the lack of variability in the first study could have played some role here. The fact that higher frequency words are usually polysemic or unspecific could also have caused these results in the latter (Crossley et al., 2007).

For its part, results of Study 2.2 regarding the familiarity of connectives, which is taken for a surface measure, are quite more complex. This time no main effect of this variable was observed, although an interaction with the kind of semantic relation stablished by the connective did come out. These results exposed that surface simplification measures do not have the direct impact on literal comprehension suggested by some simplification guidelines (e.g. use words that are currently in use and have a high frequency [Anula, Fernández-Lagunilla, Belinchón, Revilla, & Heras, 2006]).

Other measures taken a priori as surface ones (e.g. text length), turned out to have an impact on inferential but not on literal comprehension. For instance, the number of sentences in Study 1 resulted to be a negative predictor, that is to say, the higher the sentence density, the lower ability of the students with ID to identify the relations among them.

It is important to remark that the text designers followed the motto of *one idea per sentence* to create the texts in Study 1. Thus, besides the motivational effect of text length we can assert that the shortest texts were also conceptually simpler. This feature could have influenced the participants' ability to integrate ideas. From an applied point of view, we could then recommend the use of shorter texts for this kind of readers, even though this practice would also result in conceptually poorer texts.

Finally, it becomes necessary to mention the numerous researchers, who stand against the use of these surface measures (e.g. number of sentences, average number of words per sentence, Inflesz index or word frequency) to assess text complexity (Crossley et al., 2007; Davison & Kantor, 1982; Kantor & Davison, 1981). These among other authors, argue that the isolated use of surface measures to evaluate legibility and to adapt texts to an easy-to-read format overlook the adaptation at a cohesive level and is, hence, a biased praxis. As we will see in the next section, cohesive measures would not guarantee the full comprehension of the adapted texts either.

6.2.2. Effectiveness of the adaptation at the cohesive level

With regard to cohesive measures, we analysed the role of connective devices and pronouns of two different types. As for the former, in the Study 1 we obtained a negative correlation between the number of connectives and the literal comprehension, on the one hand, and the number of connectives and the number of sentences, on the other. In other words, longer texts also contained a greater amount of connectives, plus they were poorly comprehended at the literal level in comparison with the shorter texts with less connectives. Taking this correlational design as starting point it was not possible to isolate the effect of the presence of connectives. Thus, with the aim of assessing the effect of connectives independently, in the Studies 2.1 and 2.2 we set the focus on other various factors related to them. Text length and other fixed factors remained changeless while others were manipulated: presence/ absence of connectives, familiarity, and kind of semantic relation involved.

The sheer presence of connectives in our Study 2.1 did not enhance comprehension at any level. Hence, we resolved to further investigate the kinds of interclausal relations in order to find out whether the order of acquisition and the difficulties found were similar in people with ID and people with typical development (TD). When we compared the results of the group with ID with those of the reading level-matched group, qualitative differences arose: the types of connectives processed with a greater difficulty were different in both groups. On the one hand, this proves that two groups matched at their reading comprehension level do not necessarily have a similar linguistic knowledge. On the other hand, we can assure that the processing differences in the group with ID must be the product of a different pattern of acquisition rather than a mere delay of learning, although there are many classifications of connectives (Real Academia Española, 2009;

Garachana & Hilferty, 2011; Knott & Sanders, 1998), and no agreement on what the actual order of acquisition is. It is generally assumed that additive connectives (and) are acquired before causals (because), and positives (and, because) before negatives (but, although) (Bloom, 1991; Evers-Vermeul, 2005; Spooren, 1997, Pander Maat & Sanders, 2006). However, the results obtained by our children and adults without comprehension difficulties do not fit into this pattern either. As we explained in the discussion of the Study 2.2, some methodological issues could have influenced these results in an unexpected way (e.g. the use of the medium familiarity connective “antes de”, that is, “before”, due to the lack of temporal connectives of low familiarity).

The last cohesive device we have tested are co-references. In our first study we observed that the number of co-references (taken as noun repetitions, argument repetitions or stem repetitions) was the best predictor for literal comprehension. Nonetheless, the relation between both variables was inverse, thus contradicting the reviewed literature. This relation could be due either to the fact that repetitions lengthen the texts (and text length correlates negatively with literal comprehension), or to the increased redundancy that appears as a consequence of this manipulation.

The lack of variability in the Study 1 made impossible the measurement of the most studied type of co-reference in the framework of reading comprehension: pronominal anaphoras. For this reason we designed the Studies 3.1 and 3.2, where both morphosyntactic and semantic strategies were needed to resolve the anaphoras presented. Additionally, by means of the inconsistencies paradigm, in Study 3.2 we would assess not only the processing, but also the explicit detection of ambiguities related to the resolution of pronominal anaphoras.

The resolution of morphosyntactic anaphoras requires the only use of surface clues based on a morphological analysis of the pronoun and/or the possible antecedents. However, even though these are considered surface clues, a ground knowledge of the gender and number agreement grammar rules. For this reason, only the most skilled readers were expected to take advantage of this information by default. By its part, resolving semantic anaphoras demands the verification of all antecedent candidates in order to find out which one fits in the context. Obviously the previous knowledge necessary to resolve this kind of anaphora is not as specific as the one needed to resolve the morphosyntactic. As a matter of fact, only a general knowledge of the nature rules of the world is needed (e.g. knowing that the air of a fan usually cannot make a baby fly, so the most plausible antecedent for “lo” [masc.], that is, “it”, in such a context must be “folio” [masc.], “paper”). This second strategy may be more accessible when the morphosyntactic clues are not available, reason why our students with ID seem to have used it by default in both Studies 3.1 and 3.2. This fact has clearly driven them to worse results in terms of accuracy, and response and processing times, when the required strategy was the morphosyntactic. In other words, the inappropriate use of the semantic strategy results in a delay on the resolution of the anaphora.

The reading processes monitoring necessary to detect inconsistent information in texts failed in the children as well as in the students with ID. As expected, there were some significant differences between these two groups and the undergraduates in what to correct answers and inconsistency detection refers. Response times (Study 3.1) and accuracy in the objective questions (Study 3.2) reflect that the undergraduates did leverage semantic strategies to resolve the anaphoras, even though they consciously attributed the ambiguity to morphosyntactic more frequently than to semantic issues.

Finally, eye tracking data reflect strategies differing qualitatively among the group of students with ID and the two control groups: undergraduates seem to reread the text more thoroughly than the children and the youngsters with ID, also when there is no inconsistence in it. Furthermore, rereading times of consistent texts were higher in both control groups than in the experimental. In conclusion, all groups appear not to vary their reading strategy in the presence (or absence) of ambiguities affecting the pronominal co-reference. The time and effort devoted to rereading the text, however, are greater in the control groups. These results together with the poor explicit detection described in the previous paragraph lead us to conclude that the metacognitive resources our experimental group possess are indeed limited and different to the control groups'. In the framework of the comprehension-age match design (Cain et al., 2000), these results would allow the identification of reading processes monitoring as a possible cause of the reading difficulties in people with ID. This aspect would be liable to be further investigated in future longitudinal and training research designs.

Regarding text simplification, we must suggest the avoidance of morphological ambiguities, for readers with ID do not exploit sufficiently this kind of cues. In the light of our results and the literature reviewed, a future line of research might be the assessment of the explicitness of the anaphoras under the "Explicitness Principle" (Gernsbacher, 1989, 1990). This theory would predict a better performance (for instance in a sentence verification task) when the co-reference is stablished by repetition, for in this case the antecedent and the anaphor share 100% of their features. This explicitness would be especially beneficial for readers with difficulties, whose grammar and semantic skills applied to the resolution of pronouns are not satisfactory (Ehrlich et al., 1999; Shapiro & Milkes, 2004).

Alternatively, an improvement of the monitoring skills of readers with ID could be approached by means of other text modifications such as “global coherence” (Arfé et al., 2014; Arfé, Oakhill, Pianta, & Alrifai, 2012). O’Connor and Klein (2004), for instance, tested this kind of elaborations on readers with Autistic Spectrum Disorders (ASD) by underlining the pronominal anaphors in the texts presented. After each trial, the reader was asked to choose one out of three possible antecedents for the anaphor underlined, thus forcing him to monitor his own comprehension all along the reading. This manipulation improved the text comprehension notably in comparison with other elaborations, such as answering questions prior to the reading, or the control condition that presented a non-elaborated original text. This result suggest a means of improvement of the text monitoring skills that could be tested with readers with ID in the future.

7 Discusión general y conclusiones

7. Discusión general y conclusiones

La presente tesis doctoral ha tratado de profundizar en el uso de la adaptación y la simplificación de textos como medio facilitador de la comprensión lectora en jóvenes con DI. Como hemos apuntado en las introducciones de los tres artículos aquí compendiados, la falta de consenso entre los antecedentes en la literatura, así como la escasez de estudios centrados en la validación empírica de pautas de adaptación y simplificación de textos propuestas por organismos internacionales como la IFLA, especialmente para el colectivo de lectores con DI, requerían un trabajo de estas características.

Así, a lo largo de tres artículos (5 Estudios) se ha tratado de responder a dos cuestiones principales:

1) ¿Mejoran los textos simplificados la comprensión lectora en jóvenes estudiantes con DI de límite a moderada? Las medidas de comprensión lectora empleadas han sido diseñadas sobre la base del Modelo de Construcción-Integración de Kintsch (1998), distinguiéndose entre comprensión literal (superficial) y comprensión inferencial (profunda).
2) De ser así, ¿qué variables Psicolingüísticas intervienen en dicha mejora? De nuevo con el objetivo de guiar teóricamente nuestras predicciones y operativizar las medidas empleadas en la adaptación y simplificación de textos, hemos distinguido entre la adaptación a nivel superficial y a nivel de cohesión siguiendo la clasificación propuesta por Crossley, Dufty, McCarthy y McNamara (2007); Crossley, Greenfield y McNamara (2008) y Crossley y cols. (2007). A grandes rasgos, las predicciones apuntaban a un mayor impacto de las adaptaciones superficiales sobre la comprensión

literal, por un lado, y a un mayor impacto de las adaptaciones a nivel de cohesión sobre la comprensión inferencial por otro.

En esta sección recapitularemos los resultados obtenidos en los cinco Estudios tratando de responder a estas preguntas y reflexionaremos sobre las conclusiones e implicaciones de éstos desde un punto de vista aplicado.

7.1. Efecto de la adaptación de textos en la comprensión lectora de jóvenes con DI

A lo largo de este trabajo hemos podido constatar dos fenómenos. Por un lado, la población de jóvenes lectores con DI de límite a moderada con la que hemos trabajado tiene un nivel de comprensión lectora por debajo de su edad cronológica, comparable al nivel de estudiantes de 2º y 3º curso de Educación Primaria.

Por otro lado, la adaptación de textos ha mejorado su comprensión de los textos empleados, principalmente a nivel literal. En el Estudio 1 hemos podido comprobar que los textos periodísticos adaptados siguiendo las pautas de la IFLA hacen posible tanto la comprensión literal como inferencial con una tasa de respuestas correctas por encima del 70% en ambos casos. No obstante, la diferencia entre la comprensión literal y la inferencial es significativa, resultando en el segundo caso inferior que en el primero. Este último resultado fue replicado en el Estudio 2.1, en el cual observamos que la comprensión literal es significativamente mayor que la inferencial, aunque no obtuvimos efecto de la simplificación. En los Estudios 2.2, 3.1 y 3.2 no utilizamos textos adaptados, sino que evaluamos el procesamiento de diferentes variables de cohesión observando cómo afectan a la comprensión

lectora de los jóvenes con DI. En los siguientes apartados desarrollaremos este último aspecto.

7.2. Medidas lingüísticas que modulan/ predicen el efecto de la adaptación.

7.2.1. Efectividad de la adaptación a nivel superficial

Hemos observado a lo largo de los tres primeros Estudios de esta tesis que las pautas de adaptación basadas en medidas superficiales tales como la frecuencia léxica (Estudios 1 y 2.1), la familiaridad (Estudio 2.2) o el índice de legibilidad *Inflesz* (Estudio 1) no dan lugar a los efectos principales de facilitación predichos por los modelos y la literatura empírica precedente. Por ejemplo, a pesar de que hay antecedentes que han obtenido resultados positivos en el procesamiento y la comprensión al aumentar la frecuencia léxica y/o la familiaridad (Haberlandt y Graesser, 1985; Just y Carpenter, 1980; Rayner y Duffy, 1986; Crosson y cols., 2008), nuestra experiencia empírica no respalda esta mejora. El aumento de la frecuencia léxica no muestra efectos de facilitación en los Estudios 1 y 2.1 si bien, como apuntamos en los artículos respectivos, esto puede deberse a un efecto techo o de la falta de variabilidad en el primer caso. En el segundo caso, el hecho de que las palabras de mayor frecuencia suelan ser polisémicas e inespecíficas en español podría estar jugando un papel importante (Crossley y cols., 2007). Por otra parte, los resultados del Estudio 2.2 en cuanto a la familiaridad de los conectores, la cual es considerada una medida superficial, son más complejos. En esta ocasión no se observa un efecto principal de dicha medida, pero sí de interacción con el tipo de relación semántica que establece el conector. Estos resultados evidencian que las medidas superficiales de simplificación no tienen el

efecto directo en la comprensión literal que puede desprenderse de pautas de simplificación como “utilizar palabras que sean de uso actual y de frecuencia alta” (Anula, Fernández-Lagunilla, Belinchón, Revilla, y Heras, 2006).

Nos encontramos también con que otras medidas que a priori consideramos superficiales, tales como la longitud del texto, tuvieron efecto sobre la comprensión inferencial pero no sobre la literal. Por ejemplo, en el Estudio 1 el número de oraciones resultó ser un predictor negativo, es decir, una mayor densidad oracional resulta en una habilidad más pobre por parte de los estudiantes con DI para encontrar las relaciones que se establecen entre ellas.

Aparte del efecto motivacional, hay que tener en cuenta que, en el caso de los textos utilizados en el Estudio 1, los diseñadores siguieron la máxima de “una idea por oración”. Podemos afirmar entonces que los textos más cortos eran también conceptualmente más simples, lo que afectaría directamente a la capacidad de integrar ideas de los participantes. Desde un punto de vista aplicado, este resultado nos haría recomendar el uso de textos cortos para este tipo de lectores, si bien hay que matizar que esta práctica daría lugar a textos con una menor riqueza conceptual.

Por otro lado es importante señalar que un gran número de investigadores critican el uso de estas medidas superficiales (número de oraciones, número de palabras por oración, índice *Inflesz*, que es un combinación de las dos anteriores, o la frecuencia léxica) para la evaluación de la complejidad textual (Crossley y cols., 2007; Davison y Kantor, 1982; Kantor y Davison, 1981). Estos autores, entre otros, argumentan que el uso exclusivo de medidas superficiales para evaluar la legibilidad y adaptar textos a LF pierde de vista la adaptación a nivel cohesivo y que, por tanto, es una práctica sesgada. Como veremos en la siguiente sección,

en nuestro caso tampoco las medidas de cohesión garantizarían la comprensión total de los textos adaptados.

7.2.2. Efectividad de la adaptación a nivel cohesivo

Como medidas de cohesión analizamos el papel de los conectores discursivos y los pronombres de dos clases diferentes. Con respecto a los primeros, en el Estudio 1 observamos que el número de conectores correlacionaba negativamente con la comprensión literal, pero al mismo tiempo correlacionaba con el número de oraciones, esto es, en los textos más largos había también más conectores, y dichos textos se comprendían peor a nivel literal que los textos cortos con pocos conectores. Por tanto, a partir de este diseño correlacional no era posible aislar el efecto de la presencia de conectores. De esta manera, con el objetivo de comprobar el efecto de los conectores de forma independiente, en los Estudios 2.1 y 2.2 se manipularon directamente varios factores relacionados con los mismos manteniendo la extensión del textos y otros factores fijos: la presencia/ ausencia de conectores, y la familiaridad y tipo de relación semántica que estos establecen.

La mera presencia de los conectores que utilizamos en el Estudio 2.1 no ayudó a mejorar la comprensión a ningún nivel, así que, como ya hemos apuntado, en el Estudio 2.2 decidimos profundizar en los tipos de relaciones interclausales para comprobar si el orden de adquisición y las dificultades que encontraban las personas con DI eran las mismas que las encontradas por personas con desarrollo típico. Al comparar los resultados del grupo con DI y de niños de edad lectora equivalente vimos que las diferencias eran cualitativas: los tipos de conectores procesados con más dificultad eran diferentes en ambos grupos de población. Por

un lado, esto demuestra que dos grupos equiparados en el nivel de comprensión lectora no tienen por qué tener los mismos conocimientos lingüísticos específicos. Por otro lado, las diferencias de procesamiento serían entonces fruto de un diferente patrón de adquisición, no de un mero retraso en el aprendizaje por parte del grupo con DI. A este respecto hay que puntualizar que existen numerosas clasificaciones de conectores discursivos (Real Academia Española, 2009; Garachana y Hilferty, 2011; Knott y Sanders, 1998) y no hay un consenso total en cuáles se adquieren antes y después. En general se acepta que los conectores aditivos (y) se adquieren antes que los causales (porque), y los positivos (y, porque) antes que los negativos (pero, aunque) (Bloom, 1991; Evers-Vermeul, 2005; Spooren, 1997, Pander Maat y Sanders, 2006). Sin embargo, nuestros resultados en niños y adultos sin dificultades de comprensión tampoco se ajustan a este patrón aunque, como hemos visto en la discusión del Estudio 2.2, hay algunos detalles metodológicos que pueden haber influido en los resultados de manera inesperada (por ejemplo, el uso del conector de familiaridad media “antes de” como conector de baja familiaridad debido a una ausencia de conectores temporales en esta categoría).

El último elemento de cohesión que hemos puesto a prueba son las correferencias. En el Estudio 1 observamos que el número de correferencias (entendidas como repetición de un nombre, repetición de un argumento o repetición de un lexema), era la variable que mejor predecía la comprensión literal. Sin embargo, la relación entre ambas variables era inversa, lo cual contradecía las conclusiones de la literatura revisada. Esta correlación podría deberse, bien al hecho de que dichas repeticiones aumentan la longitud del texto, y ésta correlaciona a su vez negativamente con la comprensión literal, o bien a un

incremento sobre la redundancia natural que aparece como consecuencia de esta manipulación.

Por falta de variabilidad, en el Estudio 1 no se pudo medir el tipo de correferencias cuyo efecto en la comprensión lectora más se ha investigado, esto es, las anáforas pronominales. Por esta razón diseñamos los Estudios 3.1 y 3.2, en los que para resolver las anáforas presentadas era necesario tanto el uso de estrategias morfosintácticas como semánticas. Adicionalmente, en el Estudio 3.2, a través del paradigma de inconsistencias, comprobaríamos no sólo el procesamiento, sino la detección explícita de ambigüedades en el texto relacionadas con la resolución de las anáforas pronominales.

Por un lado, la resolución de anáforas morfosintácticas requiere de claves superficiales y está basada en un análisis de la forma léxica del pronombre y el/los posible/s antecedente/s. Sin embargo, a pesar de considerarlas claves "superficiales", es al menos necesario un conocimiento de las reglas lingüísticas de concordancia de género y número, razón por la cual esperábamos que sólo los lectores más hábiles se valieran de esta información como estrategia preferida. Por otro lado, para llevar a cabo una estrategia semántica es necesario comprobar qué posible antecedente "encaja" en el contexto. Naturalmente, el conocimiento previo necesario no es específico como en el caso anterior, sino general sobre las características y el funcionamiento del mundo y de las cosas (p. ej. saber que un ventilador normalmente no puede hacer volar un bebé, con lo cual el antecedente de "lo" en tal contexto debe ser "folio"). Esta segunda estrategia puede resultar más accesible cuando las claves morfosintácticas no están disponibles, y probablemente por esa razón nuestros participantes con DI la han utilizado por defecto tanto en el Estudio 3.1 como en el 3.2. Naturalmente, esto ha llevado a peores resultados en cuestión de aciertos y tiempos de respuesta y procesamiento

cuando la estrategia que se requería era morfosintáctica. En otras palabras, al ser la estrategia semántica la utilizada por defecto, su uso cuando un análisis morfosintáctico es necesario resulta en una demora en la resolución de la anáfora.

Los procesos de monitorización de la comprensión necesarios para detectar la información inconsistente en el texto también fallaron tanto en el caso de los niños como en el de los jóvenes con DI. Tal como esperábamos, hubo diferencias significativas entre estos dos grupos y los jóvenes universitarios en cuanto a respuestas correctas y detección de inconsistencias. A pesar de que este último grupo atribuyó con más frecuencia la inconsistencia a una incongruencia morfosintáctica entre el pronombre y el antecedente, los datos de tiempos de respuesta (Estudio 3.1) y de las preguntas objetivas (Estudio 3.2) indican que también se valieron de estrategias semánticas para la resolución de las anáforas.

Finalmente, los datos de movimientos oculares reflejan estrategias de lectura cualitativamente diferentes entre el grupo con DI y los grupos de control: en general los universitarios releen más el texto independientemente de la existencia de incongruencias que los niños y los jóvenes con DI. Además, los tiempos de las relecturas cuando no existe inconsistencia son mayores en los grupos de control que en el experimental. En conclusión, todos los grupos parecen no variar notablemente su estrategia de lectura al enfrentarse a un texto con incongruencias en la correferencialidad pronominal. Sin embargo, el tiempo y empeño dedicado a la relectura del texto es mayor en los participantes sin DI. Estos datos junto con los anteriores que reflejaban una pobre detección explícita, nos llevan a concluir que los recursos metacognitivos de nuestro grupo experimental son limitados y cualitativamente diferentes a los de los estudiantes de similar edad cronológica y edad lectora equivalente. Siguiendo la lógica de diseño que incluye un grupo control de similar edad de comprensión lectora (Comprehension-age match

design), estos resultados nos permiten identificar la monitorización de los procesos de lectura como posible factor causal de las dificultades comprensión lectora en DI que debe ser investigado en futuros diseños longitudinales o de entrenamiento. Por lo que respecta a la simplificación de textos, podemos concluir que las ambigüedades morfológicas deben ser evitadas, pues los lectores con DI no explotan suficientemente este tipo de claves, en nuestro caso, de género. De esta manera, a la luz de nuestros resultados y de la literatura teórica precedente, una línea futura de investigación sería la puesta a prueba de la explicitud de las anáforas siguiendo el "Principio de Explicitud" de Gernsbacher (1989, 1990). Este predice una mejor ejecución de las correferencias establecidas mediante repetición, por ejemplo en una tarea de verificación léxica, ya que el antecedente y la anáfora comparten el 100% de sus características. Esta explicitud sería especialmente ventajosa en la resolución de anáforas para los lectores con dificultades, cuyas habilidades gramaticales y semánticas en la interpretación de pronombres no son adecuadas (Ehrlich y cols., 1999; Shapiro y Milkes, 2004).

Alternativamente, con el objetivo de mejorar la capacidad de monitorización de los lectores con DI durante la comprensión lectora podrían llevarse a cabo elaboraciones de los textos, otra forma de adaptación defendida, entre otros autores, por Arfé y colaboradores (Arfé y cols., 2014; Arfé, Oakhill, Pianta, y Alrifai, 2012) y a la que denominan adaptaciones de "coherencia global". Por ejemplo, O'Connor y Klein (2004) aplicaron este tipo de elaboraciones en lectores con Trastornos del Espectro Autista (TEA). La elaboración consistía en la señalización de anáforas pronominales mediante el subrayado de las mismas en el texto. Tras la aparición de cada pronombre subrayado en el texto, se pedía al lector que eligiera uno de tres posibles antecedentes que aparecían en el texto de manera que se forzaba al lector a monitorizar su comprensión durante la lectura. Esta

manipulación mejoró significativamente la comprensión del texto por parte de los lectores con TEA con respecto a otro tipo de elaboraciones como la resolución de preguntas previas a la intervención o la condición control que presentaba el texto original sin ningún tipo de elaboración. Este resultado sugiere una vía de mejora de las capacidades de monitorización del texto que podrían ponerse a prueba con lectores con DI en el futuro.

8 Limitaciones de la tesis

8. Limitaciones de la tesis

Ya hemos mencionado anteriormente el carácter novedoso de esta tesis, no sólo por la temática de la simplificación y adaptación de textos, sino también por la población con la que hemos trabajado, e incluso por la metodología. Sin embargo, como cualquier trabajo de investigación, éste posee algunas limitaciones que es necesario mencionar, la mayoría de ellas en cuanto a la metodología. En esta sección haremos algunas aclaraciones a este respecto.

Una de las dificultades con la que nos hemos encontrado y, por extensión, una limitación ha sido la falta de antecedentes en la literatura. Si bien este vacío ha propiciado el desarrollo de un trabajo novedoso como éste, el carecer de fuentes que ayuden a guiar el diseño y proyección de un estudio tan amplio obliga a hacer extensivas otras fuentes desarrolladas con poblaciones diferentes. Por ejemplo, gran parte de los estudios sobre lenguaje y DI se centran en síndromes genéticos como el Síndrome de Down, pero no tanto en DI de etiología diversa y en muchos casos desconocida. Aunque las habilidades afectadas en uno y otro caso son con frecuencia dispares, en ocasiones hemos tenido que asumir como válidos estudios realizados con dichas poblaciones por ser sus características más cercanas a las de la población objetivo de la tesis que a las personas de desarrollo típico.

Respecto a la muestra estudiada, tanto su tamaño reducido como la variabilidad en algunas de sus características podrían haber afectado a los resultados de manera no controlada. Como hemos visto en la sección *metodología y muestra*, Tabla 2, el número de individuos que conformaban inicialmente la muestra, especialmente el grupo experimental, era muy superior al que se empleó al final. Las razones por las que se fue reduciendo han sido descritas en las secciones anteriores. Debido a las características concretas que se requerían de los

participantes resultó inviable ampliarla a posteriori. En lo que respecta a las etiologías, la heterogeneidad del grupo hizo que finalmente controláramos y descartáramos sólo a aquellos participantes con diagnóstico múltiple que podían afectar de manera no deseada al desarrollo de la tarea, como por ejemplo aquellas personas con DI que mostraban déficits atencionales asociados, problemas de conducta o trastornos del espectro autista.

En cuanto a los grupos seleccionados, y entrando en las limitaciones del diseño, tendría sentido haber realizado todos los Estudios con el mismo número de grupos y tipos de población (DI, niños de edad lectora equivalente y adultos de similar edad cronológica). Sin embargo, a menudo por limitaciones de acceso a recursos, tiempo y/o espacio, esto no siempre fue posible.

Algunos de los Estudios que hemos presentado parten de ideas de otros autores. Es el caso de los Estudios 2.2, 3.1 y 3.2. En el caso del tercero, la tarea se adaptó del inglés al español. A pesar de ser equivalentes en familiaridad y contenido semántico, el procesamiento en español puede haber sido diferente, provocando que nuestros resultados no repliquen los de Crosson y cols. (2008). De manera similar, el diseño original del Estudio 3.1 (Elosúa y cols., 2009) era entre grupos y nosotros lo convertimos en intra-grupo. Este cambio podría explicar por qué los universitarios que participaron en el estudio original y en el nuestro parecen haber empleado estrategias diferentes para resolver las anáforas.

Finalmente, a la hora de interpretar la tesis en conjunto no debemos perder de vista que se ha tratado de poner a prueba el efecto sobre la simplificación de textos de un conjunto reducido de elementos lingüísticos relacionados con el texto y el lector mientras que los compendios de pautas de adaptación recomendadas por entidades de estandarización internacionales como la IFLA incluyen modificaciones en la presentación, tipografía empleada e incluso ayudas y

aclaraciones sobre el léxico, cuestiones que han quedado fuera del presente trabajo. También en el ámbito de la investigación se proponen niveles de adaptación de textos que no se han considerado en la presente tesis. Por ejemplo, como ya hemos mencionado anteriormente, lo que Arfé, Oakhill y Pianta (2014) denominan "coherencia global", esto es, elaboraciones que hacen explícita la información necesaria para entender el significado general de la historia, la secuencia de eventos o la relación entre la historia y el conocimiento previo del lector. Por tanto, las críticas a la aplicación de algunas pautas de simplificación que se desprenden de los resultados de esta tesis deben quedar circunscritas a los aspectos concretos que se han sometido a evaluación empírica no haciéndose extensivos a otros fenómenos relacionados con la adaptación de textos sobre los que la tesis no ha versado. Asimismo, de esta limitación se desprende que el trabajo futuro debe ir encaminado a ampliar la validación empírica con el colectivo de jóvenes con DI a otros niveles de adaptación y simplificación de textos.

Referencias

REFERENCIAS

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

American Psychiatric Association. (2000). *DSM-IV*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. *Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Publishing. http://doi.org/10.1176/appi.books.9780890425596.893619

Anula, A., Fernández-Lagunilla, M., Belinchón Carmona, M., Revilla Guijarro, A., & Heras Sedano, L. (2006). *Introducción a Don Quijote de la Mancha de Fácil Lectura*.

Arfé, B., Oakhill, J., & Pianta, E. (2014). The Text Simplification in TERENCE. In T. D. M. et Al. (Ed.), *Advances in Intelligent Systems and Computing* (Vol. 292, pp. 165–172). Switzerland: Springer International Publishing. http://doi.org/10.1007/978-3-319-07698-0_21

Arfé, B., Oakhill, J., Pianta, E., & Alrifai, M. (2012). *Story simplification user guide. Technical report D.2.2, TERENCE Project*.

Bloom, L. (1991). *Language development from two to three. Language development from two to three.* Retrieved from http://libproxy.tulane.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1991-98370-000&site=ehost-live&scope=site

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. http://doi.org/10.1037/a0022824

Cain, K., Oakhill, J. V., & Bryant, P. E. (2000). Investigating the causes of reading comprehension failure: the comprehension-age match design. *Reading and Writing: An Interdisciplinary Journal*, *12*, 31–40. http://doi.org/10.1023/A:1008058319399

Channell, M. M., Loveall, S. J., & Conners, F. a. (2013). Strengths and weaknesses in reading skills of youth with intellectual disabilities. *Research in Developmental Disabilities*, *34*(2), 776–87. http://doi.org/10.1016/j.ridd.2012.10.010

Consejería de Educación de la Comunidad Valenciana. DECRETO 111/2007 (2007). Comunidad Valenciana, España.

Crossley, S. a., Dufty, D. F., McCarthy, P. M., & McNamara, D. S. (2007). Toward a new readability: A mixed model approach. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 197–202.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, *42*, 475–493. http://doi.org/10.1002/j.1545-7249.2008.tb00142.x

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & Mcnamara, D. S. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, *91*(1), 15–30. http://doi.org/10.1111/j.1540-4781.2007.00507.x

Crosson, A. C., Lesaux, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, *29*(04), 603–625. http://doi.org/10.1017/S0142716408080260

Davison, A., & Kantor, R. N. (1982). On the Failure of Readability Formulas to Define Readable Texts: A Case Study from Adaptations. *Reading Research Quarterly*, *17*(2), 187. http://doi.org/10.2307/747483

De la Cruz, M. (1999). *Evaluación de la Comprensión Lectora (ECL-1)*. Madrid: TEA.

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III: Peabody test de vocabulario en imágenes*. Madrid: TEA.

Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal*, *11*, 29–63.

Elosúa, M. R., Carriedo, N., & García-Madruga, J. A. (2009). Dos nuevas pruebas de Memoria Operativa de Anáforas. *Infancia Y Aprendizaje*, *32*(1), 97–118. http://doi.org/10.1174/021037009787138185

Española, R. A. (2009). *Nueva Gramática de la Lengua Española, 2 vols.* Madrid: Espasa Libros.

Evers-Vermeul, J. (2005). *Connections between form and function of Dutch Connectives: Change and acquisition as windows on form-function relation*. Universiteit Utrecht.

Garachana, M., & Hilferty, J. (2011). Marcador discursivo. In *Diccionari de lingüística on line*.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*(2), 99–156. http://doi.org/10.1016/0010-0277(89)90001-2

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Haberlandt, K. (1982). Reader Expectations in Text Comprehension. In L. Ny & W. Kintsch (Eds.), *Language and Comprehension* (pp. 239–249). Amsterdam: North-Holland.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*. http://doi.org/10.1037/0096-3445.114.3.357

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. http://doi.org/10.1037/0033-295X.87.4.329

Kantor, R. N., & Davison, A. (1981). Categories and strategies of adaptation in children's reading materials. In A. Rubin (Ed.), *Conceptual readability: New ways to look at text, reading education report n° 31* (pp. 54–68). Washington, DC: National Institute of Education Center for the Study of Reading.

Kaufman, A. S., & Kaufman, N. L. (1997). *Test breve de inteligencia de Kaufman (K-BIT)*. Madrid: TEA.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press.

Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, *30*(2), 135–175. http://doi.org/10.1016/S0378-2166(98)00023-X

Koda, N. (2008). Connective interference and facilitation: Do connectives really facilitate the understanding of discourse? *The Annual Reports of Graduate School of Arts and Letters*, *56*, 29–42.

Lundberg, I., & Reichenberg, M. (2013). Developing Reading Comprehension Among Students With Mild Intellectual Disabilities: An Intervention Study. *Scandinavian Journal of Educational Research*, *57*(1), 89–100. http://doi.org/10.1080/00313831.2011.623179

Martínez, T., Vidal-Abarca, E., Sellés, P., & Gilabert, R. (2008). Evaluación de las estrategias y los procesos de comprensión: el test de procesos de comprensión (TPC). *Infancia Y Aprendizaje*, *31*(3), 319–332.

Millis, K. K., & Just, M. A. (1994). The Influence of Connectives on Sentence Comprehension. *Journal of Memory and Language*, *33*(1), 128–147. http://doi.org/10.1006/jmla.1994.1007

Morgan, M. F., & Moni, K. B. (2008). Meeting the challenge of limited literacy resources for adolescents and adults with intellectual disabilities. *British Journal of Special Education*, *35*(2), 92–101. http://doi.org/10.1111/j.1467-8578.2008.00378.x

Murray, J. D. (1995). Logical Connectives and Local Coherence. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of Cohesion in Text Comprehension* (pp. 107–125). Hillsdale, N.J.: Lawrence Erlbaum.

Nomura, M., Nielsen, G. S., & Tronbacke, B. (2010). *Guidelines for easy-to-read materials: Revision*. The Hague: IFLA Headquarters.

O'Connor, I. M., & Klein, P. D. (2004). Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *34*(2), 115–127. http://doi.org/10.1023/B:JADD.0000022603.44077.6b

Pander Maat, H., & Sanders, T. (2006). Connectives in Text. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics. Volume 3* (pp. 33–41). Amsterdam: Elsevier.

Quigley, S. P., & King, C. (n.d.). Reading milestones. Beaverton, OR: Dormac.

Ramos, J. L., & Cuetos, F. (1999). *PROLEC-SE: evaluación de los procesos lectores en alumnos del tercer ciclo de educación primaria y educación secundaria obligatoria*. Madrid: TEA.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. http://doi.org/10.3758/BF03197692

Rinck, M., Gámez, E., Díaz, J. M., & De Vega, M. (2003). Processing of temporal information: evidence from eye movements. *Memory & Cognition*, *31*(1), 77–86.

Shapiro, A., & Milkes, A. (2004). Skilled readers make better use of anaphora: A study of the repeated-name penalty on text comprehension. *Electronic Journal of Research in Educational Psychology*, *2*(2), 161–180.

Siddharthan, A. (2014). A survey of research on text simplification. *ITL - International Journal of Applied Linguistics*, *165*(2), 259–298. http://doi.org/10.1075/itl.165.2.06sid

Spooren, W. (1997). The processing of underspecified coherence relations. *Discourse Processes*. http://doi.org/10.1080/01638539709545010

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics*, *40*(12), 2003–2026. http://doi.org/10.1016/j.pragma.2008.04.021

Tronbacke, B. (1997). *Guidelines for Easy-to-read Materials*. (I. Headquarters, Ed.).

Van den Bos, K. P., Nakken, H., Nicolay, P. G., & van Houten, E. J. (2007). Adults with mild intellectual disabilities: can their reading comprehension ability be improved? *Journal of Intellectual Disability Research : JIDR*, *51*(Pt 11), 835–49. http://doi.org/10.1111/j.1365-2788.2006.00921.x

Van der Molen, M. J., Van Luit, J. E. H., Jongmans, M. J., & Van der Molen, M. W. (2007). Verbal working memory in children with mild intellectual disabilities. *Journal of Intellectual Disability Research*, *51*(2), 162–169. http://doi.org/10.1111/j.1365-2788.2006.00863.x

Wechsler, D. (2005). *Escala de Inteligencia Wechsler para Niños IV. (WISC-IV)*. Madrid: TEA.

Anexos

Anexo A

Estudio 1. Ejemplo de texto (noticia internacional), variables lingüísticas extraídas y preguntas de comprensión.

Hoy es el día mundial de la discapacidad

Es un día especial.

Hoy se celebra el Día Mundial de la Discapacidad. Es un día especial porque se recuerdan los derechos de las personas con discapacidad.

Un derecho de las personas con discapacidad es tener un trabajo.

Las empresas como “Manpower” buscan empleo a las personas. Desde hace cinco años también buscan trabajo a personas con discapacidad. Cada vez son más personas con discapacidad las que encuentran trabajo.

Las empresas se dan cuenta de que es bueno contratar a una persona con discapacidad.

Muchas empresas adaptan el puesto de trabajo según las necesidades de la persona que empieza a trabajar. Por ejemplo, cambian los picaportes de las puertas y el escritorio para que la persona con discapacidad pueda moverse mejor.

Se han conseguido muchas cosas con la integración laboral y todavía quedan muchas cosas por hacer.

Tabla A1

Variables extraídas del texto de ejemplo.

Variables léxicas	
Número de palabras	145
Promedio de sílabas por palabra	2.01
Promedio de frecuencia léxica	20279
Variables sintácticas	
Número de oraciones	16
Promedio de palabras por oración	9.06
Índice Flesch-Szigriszt	72.31
Grado en la escala Inflesz	Bastante fácil
Variables de cohesión	
Número de conectores (total)	5
Número de correferencias	13
Número de elipsis	4

Tabla A2

Preguntas de comprensión correspondientes al texto de ejemplo.

1.	*El día mundial de la discapacidad es especial…* (**L**)	
	Porque hay más personas con discapacidad que tiene trabajo.	○
	Porque un derecho de las personas con discapacidad es tener un trabajo.	○
	Porque se recuerdan los derechos de las personas con discapacidad.	●
2.	*¿Cómo adaptan el puesto de trabajo las empresas?* (**I**)	
	Según las necesidades de la persona que empieza a trabajar.	○
	Cambian las puertas y el escritorio.	○
	Hacen que la persona con discapacidad pueda moverse mejor.	●
3.	*Di lo que es verdad* (**I**)	
	Las empresas como Manpower se dan cuenta de que es bueno contratar a una persona con discapacidad.	○
	Manpower ayuda a las personas con discapacidad a encontrar trabajo desde hace seis años.	○
	Las empresas como Manpower ayudan a que las personas con discapacidad cumplan el derecho de tener un trabajo.	●

Nota. Entre paréntesis se explicita el tipo de pregunta: L = literal; I = inferencial. Se marca con un punto negro la respuesta correcta.

Anexo B

Estudio 2.1. Ejemplo de noticia no adaptada, con las tres correspondientes adaptaciones (frecuencia léxica, conectividad y versión combinada) y cuestionario de comprensión.

Versión no adaptada

Río de Janeiro gana, Madrid dice adiós al sueño olímpico

Madrid ha caído en la final ante Río de Janeiro y ha rozado el sueño olímpico. Río de Janeiro ha obtenido 66 votos y Madrid 32. Chicago, una de las máximas favoritas, ha sido excluida en la primera votación del Comité Olímpico Internacional. Madrid ha ganado la primera vuelta con 28 votos frente a los 26 de Río, los 22 de Tokio y los 18 de la ciudad de Obama que ha caído ante la sorpresa general.

El Comité Olímpico Internacional ha descartado a Chicago y ha dejado que continúen en carrera Tokio, Madrid y Río de Janeiro. Una bomba que nadie esperaba. Todas las apuestas señalaban que era la predilecta tras la asistencia del presidente de Estados Unidos, Barack Obama, a la presentación del proyecto de Chicago ante el Comité Olímpico Internacional. Tras Chicago, ha caído Tokio en la segunda vuelta. Río obtuvo 46 votos, Madrid 29 y Tokio 20. Los españoles y brasileños quedaban finalistas.

Los centenares de madrileños congregados en la Plaza de Oriente botaron de alegría al escuchar que Chicago y Tokio dejaban de estar en la carrera olímpica, y ver "más cerca" el sueño olímpico de Madrid. La alegría no duró mucho, una hora, lo que tardó el Comité Olímpico Internacional en anunciar la triunfadora. Abrazos y caras tristes en Madrid tras la elección de Río de Janeiro.

Adaptación de la frecuencia léxica

Río de Janeiro gana, Madrid dice adiós al sueño olímpico

Madrid ha caído en la final ante Río de Janeiro y ha tocado el sueño olímpico. Río de Janeiro ha conseguido 66 votos y Madrid 32. Chicago, una de las mayores favoritas, ha sido eliminada en la primera vuelta del Comité Olímpico Internacional. Madrid ha ganado la primera vuelta con 28 votos frente a los 26 de Río, los 22 de Tokio y los 18 de la ciudad de Obama que ha caído ante la sorpresa general..

El Comité Olímpico Internacional ha eliminado a Chicago y ha dejado que sigan en carrera Tokio, Madrid y Río de Janeiro. Un resultado que nadie esperaba. Todas las apuestas señalaban que era la favorita tras la participación del presidente de Estados Unidos, Barack Obama, en la presentación del proyecto de Chicago ante el Comité Olímpico Internacional. Tras Chicago, ha caído Tokio en la segunda vuelta. Río consiguió 46 votos, Madrid 29 y Tokio 20. Los españoles y brasileños quedaban finalistas.

Los cientos de madrileños reunidos en la Plaza de Oriente botaron de alegría al escuchar que Chicago y Tokio dejaban de estar en la carrera olímpica, y ver "más cerca" el sueño olímpico de Madrid. La alegría no duró mucho, una hora, lo que tardó el Comité Olímpico Internacional en anunciar la elegida. Abrazos y caras tristes en Madrid tras la elección de Río de Janeiro.

Adaptación de la conectividad

Río de Janeiro gana, Madrid dice adiós al sueño olímpico

Madrid ha caído en la final ante Río de Janeiro después de que ha rozado el sueño olímpico ya que Río de Janeiro ha obtenido 66 votos y Madrid 32. Chicago, una de las máximas favoritas, ha sido excluida en la primera votación del Comité Olímpico Internacional. Sin embargo, Madrid ha ganado la primera vuelta con 28 votos frente a los 26 de Río, los 22 de Tokio y los 18 de la ciudad de Obama que ha caído ante la sorpresa general.

El Comité Olímpico Internacional ha descartado a Chicago y ha dejado que continúen en carrera Tokio, Madrid y Río de Janeiro. O sea, una bomba que nadie esperaba ya que todas las apuestas señalaban que era la predilecta tras la asistencia del presidente de Estados Unidos, Barack Obama, a la presentación del proyecto de Chicago ante el Comité Olímpico Internacional. Tras Chicago, ha caído Tokio en la segunda vuelta ya que Río obtuvo 46 votos, Madrid 29 y Tokio 20. Es decir, los españoles y brasileños quedaban finalistas.

Los centenares de madrileños congregados en la Plaza de Oriente botaron de alegría al escuchar que Chicago y Tokio dejaban de estar en la carrera olímpica, y ver "más cerca" el sueño olímpico de Madrid. Pero la alegría no duró mucho, una hora, es decir, lo que tardó el Comité Olímpico Internacional en anunciar la triunfadora. Así que abrazos y caras tristes en Madrid tras la elección de Río de Janeiro.

Versión combinada

Río de Janeiro gana, Madrid dice adiós al sueño olímpico

Madrid ha caído en la final ante Río de Janeiro después de que ha tocado el sueño olímpico ya que Río de Janeiro ha conseguido 66 votos y Madrid 32. Chicago, una de las mayores favoritas, ha sido eliminada en la primera vuelta del Comité Olímpico Internacional. Sin embargo, Madrid ha ganado la primera vuelta con 28 votos frente a los 26 de Río, los 22 de Tokio y los 18 de la ciudad de Obama que ha caído ante la sorpresa general.

El Comité Olímpico Internacional ha eliminado a Chicago y ha dejado que sigan en carrera Tokio, Madrid y Río de Janeiro. O sea, una resultado que nadie esperaba ya que todas las apuestas señalaban que era la favorita tras la participación del presidente de Estados Unidos, Barack Obama, a la presentación del proyecto de Chicago ante el Comité Olímpico Internacional. Tras Chicago, ha caído Tokio en la segunda vuelta ya que Río consiguió 46 votos, Madrid 29 y Tokio 20. Es decir, los españoles y brasileños quedaban finalistas.

Los cientos de madrileños reunidos en la Plaza de Oriente botaron de alegría al escuchar que Chicago y Tokio dejaban de estar en la carrera olímpica, y ver "más cerca" el sueño olímpico de Madrid. Pero la alegría no duró mucho, una hora, es decir, lo que tardó el Comité Olímpico Internacional en anunciar la elegida. Así que abrazos y caras tristes en Madrid tras la elección de Río de Janeiro.

Tabla B1

Preguntas de comprensión correspondientes los textos de ejemplo.

1.	*¿Quién ha perdido contra Río de Janeiro en la final del concurso para ser sede de las Olimpiadas?* (**L**)	
	Madrid	●
	Brasil	○
	Chicago	○
2.	*¿Qué ciudad favorita fue eliminada en la primera vuelta?* (**L**)	
	Madrid	○
	Brasil	○
	Chicago	●
3.	*¿Qué presidente ha participado en la presentación del proyecto de su ciudad ante el Comité Olímpico Internacional?* (**L**)	
	José Luis Rodríguez Zapatero	○
	El presidente de Brasil	○
	Barack Obama	●

4. *La noticia dice que Madrid ha tocado el sueño olímpico porque…* (**I**)

Iba ganando a otras ciudades al principio del concurso pero ha perdido en la final.	●
Ha ganado a Río de Janeiro en la segunda vuelta.	○
Ha ganado a Tokio en la primera vuelta.	○

5. *¿Cuántos votos ha tenido Madrid en la tercera vuelta del concurso para ser sede de las Olimpiadas?* (**I**)

29	○
32	●
66	○

6. *¿Por qué nadie esperaba que Chicago quedara eliminada?* (**I**)

Porque el Presidente Obama la había apoyado.	●
Porque Comité Olímpico Internacional la había apoyado.	○
Porque es una ciudad norteamericana.	○

7. *¿Por qué Españoles y Brasileños quedaron para la final?* (**I**)

Porque Madrid y Río de Janeiro sacaron más votos que Tokio.	●
Porque Madrid y Tokio, sacaron más votos que Río de Janeiro.	○
Porque Madrid y Río son más alegres que las demás ciudades.	○

8. *¿Estaban alegres los madrileños en la Plaza de Oriente?* (**I**)

Sólo hasta que supieron que los brasileños habían ganado.	●
Sí, durante más de una hora.	○
Sí, porque habían ganado el concurso.	○

9. *¿Cómo han quedado Chicago y Madrid en la primera vuelta del concurso?* (**I**)

Igual puntuación para las dos ciudades F	○
Chicago eliminado y Madrid no C	○
Chicago y Madrid eliminados por Brasil F	●

Nota. Entre paréntesis se explicita el tipo de pregunta: L = literal; I = inferencial. Se marca con un punto negro la respuesta correcta.

Anexo C

Estudio 2.2. Ejemplo de estímulo con las opciones de respuesta correspondientes.

Tabla C1

Ejemplo de la categoría de conectores contrastivos.

Baja familiaridad	
El animal preferido de Susana es el perro._____, Carolina prefiere los gatos	
por el contrario	●
por lo tanto	○
pese a	○
Alta familiaridad	
Si no comes el primer plato _____ comerás el postre	
tampoco	●
porque	○
en cambio	○

Nota. Se marca con un punto negro la respuesta correcta.

Anexo D

Estudio 3.1. Ejemplo de estímulo con las opciones de respuesta correspondientes.

Tabla D1

Ejemplo para el pronombre masculino “lo”.

Anáfora morfosintáctica	
*Ricardo **lo** eligió para el equipo porque tenía muy buenas referencias*	
jugador	●
nadadora	○
Anáfora semántica	
*Ricardo **lo** eligió para el equipo porque tenía muy buenas referencias*	
jugador	●
coche	○

Nota. Se marca con un punto negro la respuesta correcta.

Anexo E

Estudio 3.2. . Ejemplo de texto y preguntas de comprensión.

Tabla E1

Texto con pronombre masculino.

0	"Deporte con canciones de Coti"
1.	Jorge y Clara quedaron para ir a correr al parque ayer.
2a.	Consistente:
	Jorge dijo a Clara un truco para no cansarse mucho.
2a.	Inconsistente:
	Clara dijo a Jorge un truco para no cansarse mucho.
3.	El truco era calentar y escuchar las canciones de Coti a la vez.
4.	Así que Clara y Jorge siguieron este truco y no se cansaron al correr.
5.	Consistente:
	Él tenía muchos trucos útiles.
6.	La próxima semana volverán a hacer lo mismo.

Nota. A cada participante se le presentó bien 2a o 2b.

Tabla E2

Preguntas de comprensión correspondientes al texto de ejemplo.

1.	*¿Para qué quedaron Clara y Jorge?*	
	Para correr	●
	Para escuchar canciones	○
	Para hablar	○
2.	*¿De quién era el truco para no cansarse corriendo?*	
	Jorge	●
	Clara	○
	Coti	○

3.	*¿Quién tenía muchos trucos útiles?*	
	Jorge	●
	Clara	○
	Coti	○
4.	*¿Hay algo raro en el texto? ¿Qué?*	RA
5.	*¿Qué es lo más difícil del texto?*	RA

Nota. Se marca con un punto negro la respuesta correcta. RA = respuesta abierta.