Simplificación p/subtitulado

- Aprendizaje: copiar palabra, reemplazar palabra, borrar palabra, insertar (ignorado en el modelo)
- Se alinea la transcripción y los subtítulos a nivel palabra, se quedan con aquellos pares que comparten al menos la mitad de las palabras
- Se aplica un método de aprendizaje basado en "memoria" que finalmente no funciona demasiado bien

Simplificación de textos de alta complejidad

- Simplificación de patentes en PATExpert (Bouayad-Agha et al. 2009)

An optical disk drive comprising: a laser light source for emitting a laser beam; an optical system for conversing the laser beam from the laser light source on a signal plane of optical disk on which signal marks are formed and for transmitting the light reflected from the signal plane; one or more optical components, arranged in the optical path between the laser light source and the optical disk, for making the distribution of the laser beam converged by the conversing means located on a ring belt just after the passage of an aperture plane of the optical system; a detection means for detecting the light reflected from the optical disk; and a signal processing circuit for generating a secondary differential signal by differentiating the signals detected by the detection means and for detecting the edge positions of the signal marks by comparing the secondary differential signal with a detection level. (ORACION ORIGINAL)

An optical disk drive comprises a laser light source, an optical system, a detection means, and a signal processing circuit.
The laser light source emits a laser beam.
The optical system converses the laser beam from the laser light source on a signal plane of optical disk.
On the latter, signal marks are formed.
The optical system also transmits the light reflected from the signal plane.

The signal processing circuit generates a secondary differential signal.
To do so, it differentiates the signals detected by the detection means.
It also detects the edge positions of the signal mark.
To do so, it compares the secondary differential signal with a detection level. (ORACIONES SIMPLIFICADAS)

Trabajos conexos

- El proyecto FIRST "Flexible Interactive Reading Support Tool" (proyecto FP7) (Mitkov, 2012)
  - Simplificación de textos para colectivos con autismo = problemas con oraciones complejas, ambigüedad, lenguaje figurativo, etc.
  - Simplificación a nivel de los textos, herramientas de navegación, agregado de pictogramas, etc.
  - Simplificación para el usuario autista pero también para los allegados al usuario

Trabajos conexos

- Simplificación para la dislexia (Rello et al. 2012)
  - se trabaja en detección de palabras consideradas complicadas para disléxicos
  - se realiza un estudio con tecnología de eye-track para verificar la influencia de frecuencia y longitud de palabra en el entendimiento
  - se esta desarrollando un método de substitución por palabras que sean mas apropiadas

Trabajos conexos

- Simplificación de expresiones numéricas (Bautista et al. 2012) y las adapta para el español (en curso)
  - estudia como hacer mas accesibles las expresiones numéricas utilizando por ejemplo: redondeo o agregando/modificando modificadores

Original	Simplificaciones
Cerca de 1,9 millones de personas asistieron al concierto	Casi 2 millones de personas asistieron al concierto
Sólo se ha vendido un cuarto de las entradas	Sólo se ha vendido ¼ de las entradas
Uno de cada cuatro niños hablan chino	1 de cada 4 niños hablan chino
Asistieron un 57% de la clase	Asistieron mas de la mitad de la clase
Aprobaron el 98%	Aprobaron casi todos

© Horacio Saggion - 2012

El proyecto Simplext

- Desarrollar un sistema de simplificación de textos para personas con discapacidad cognitiva. Tiene los auspicios del Ministerio de Industria, Turismo y Comercio. (Saggion et al, 2011)
- Una de las primeras aplicaciones de simplificación de textos para el español.
- Objetivo tecnológico
  - Desarrollo de una sistema ubicuo de simplificación
  - Que el contenido simplificado pueda ser consumido en diferentes dispositivos
- Objetivos científicos
  - Desarrollo de recursos para la simplificación en español
  - Desarrollo de métodos y algoritmos para simplificar contenido
- Objetivo social: simplificar el acceso a la información por un colectivo con discapacidad

© Horacio Saggion - 2012

El proyecto Simplext

- El español es la lengua oficial de solo un país en Europa: España
- Sin embargo es la lengua oficial en otros 20 países
- Es hablado por más de 500 millones de personas como lengua materna o segunda lengua
- Es la tercera lengua con mas presencia en Internet luego del inglés y el chino

© Horacio Saggion - 2012

El consorcio Simplext

- Compañias tecnológicas: Abada, Ariadna, SADE Consulting, Technosite, Tilo
- Usuarios: Fundación Prodis
- Proveedores de datos: Servimedia
- Legal: CE Consulting
- Expertos en adaptación manual de textos: Universidad Autónoma de Madrid (DILES)
- Procesamiento de lenguaje natural: Universitat Pompeu Fabra
  - **Horacio Saggion**
  - **Stefan Bott**
  - **Biljana Drndarevic**
  - **Simon Mille**

© Horacio Saggion - 2012

Simplificación en Simplext

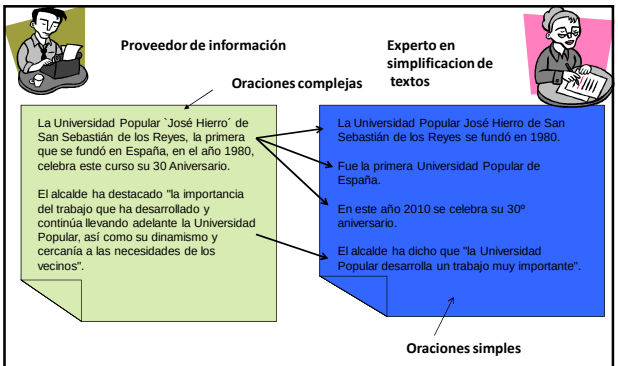
- Simplificación léxica (Drndarevic B, Saggion H., 2012a)
  - Escoger un vocabulario accesible
- Reducir la complejidad de las oraciones (Bott S, Saggion H., 2012)
  - Reducir su longitud y su estructura
- Reducir contenido (Drndarevic B, Saggion H., 2012b)
  - Solo conservar lo que es mas importante
- Clarificar (Drndarevic B, Saggion H., 2012b)
  - Cualquier termino que así lo requiera

Metodología en Simplext

- Desarrollo de recursos de lenguajes y herramientas computacionales
  - Primer corpus para el estudio de la simplificación de textos en español = 200 textos y sus simplificaciones asociadas
  - Desarrollo de un léxico para la simplificación basado en el corpus de referencia de la Real Academia
  - Desarrollo de herramientas de alineación de oraciones y herramientas para preparar el corpus
- Análisis del corpus
  - Identificación de operaciones de simplificación/transformaciones
- Desarrollo de software
  - Motor de simplificación de textos
- Integración del software en varias aplicaciones
- Evaluación con usuarios

Simplificación manual

- El método de simplificación manual es desarrollado por el grupo DILES de la UAM (Anula 2008, 2009) se interesa por dos variables en el texto: vocabulario y sintaxis
  - Vocabulario
    - frecuencia, densidad léxica
  - Sintaxis
    - Textos complejos tienen oraciones largas, coordinaciones, subordinaciones, estructuras recursivas, complementos no obligatorios
    - Un texto simple se obtiene evitando estas construcciones : usar oraciones cortas, sin estructuras de subordinación, etc.

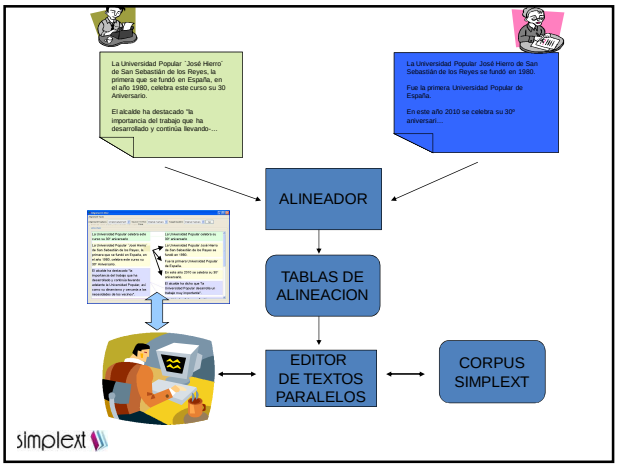


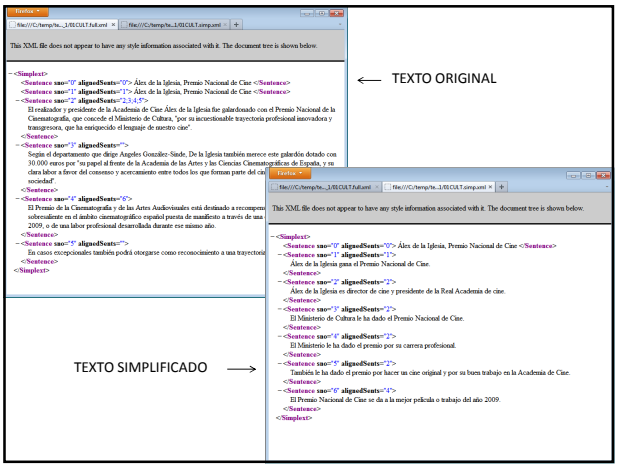
Alineación de textos

- Objetivo saber que oraciones en el texto simplificado corresponden a oraciones en el texto original
- Hipótesis de trabajo (Bott & Saggion, 2011)
  - Nivel léxico: hay muchas palabras en común entre la oración original y su simplificación (diferente de traducción automática)
  - Nivel textual: el orden en que se presenta la información en la simplificación es similar al orden en que la información se presenta en el original (similar a traducción automática)
- Se modela la "probabilidad" de que dos oraciones estén alineadas usando una medida de similitud léxica
- Un proceso de Markov intenta predecir las posiciones del original de las cuales se emitieron las simplificadas

Alineación de textos

- Luego de desarrollado el algoritmo se aplico a una muestra de documentos a simplificar (110 oraciones en textos originales, 145 oraciones simples)
- Se alinearon manualmente las oraciones
- Se desarrollo un sistema "baseline" basado en similitud léxica
- Se midió la cobertura y la precisión del algoritmo comparando con la alineación manual
- Se alcanzo una precisión de 82% y una cobertura de casi un 81%, mejor que el baseline y que lo reportado en la literatura
- El algoritmo comete errores que es necesario corregir





Estudio en simplificación lexica

- Bases para el desarrollo de un módulo de simplificación léxica
- Se observó que alrededor de 18% de las operaciones de simplificación son cambios léxicos que pueden ser:
  - substitución de una palabra por un "sinónimo"
    - flora => plantas
  - simplificación de expresiones numéricas y fechas
    - 100,000 personas => 100 mil personas; 2010 => le año 2010
  - eliminación de nominalizaciones
    - reapertura => abrir, colaboración => colaborar
  - reformulación de expresiones idiomáticas y colocaciones
    - acoger muestra => hacer exposición
  - reformulación de expresiones metafóricas
    - gigante de la escena => gran actor, precios se disparan => las cosas están muy caras

Análisis de los datos

	Original	Simple
Total de palabras	6595	3912
Total de oraciones	246	324
promedio oración/texto	6.64	8.75
promedio palabra/texto	26.8	12.07
promedio long. palabra	5.44	5.07

- textos simples cerca de 40% menos palabras, pero casi un 50% mas de oraciones
- oraciones en el original se dividen en varias oraciones
- algunas oraciones son directamente eliminadas

Longitud de palabra

- Se analizaron palabras de longitud: 1, 2, 3…20 , longitud 21-30, longitud 31-40, longitud 40+
- Notar que las palabras largas son generalmente nombres
  - palabras de longitud 2 son las mas comunes en textos originales y simplificados
  - palabras de longitud media (3-7) son mas comunes en textos simples
  - palabras mas largas mas comunes en textos originales
  - hay palabras de longitud grande (18, 21-30) mas comunes en textos simples y esto de debe a que los nombres se repiten en su forma completa en los textos simplificados

Análisis de la frecuencia

- Utilizamos un diccionario de frecuencias creado a partir del corpus de la Real Academia del Español
- Índice de frecuencia (FI): 1-6 (6: muy frecuente, 1: infrecuente)
- Consideramos que palabras que no están en el diccionario tienen FI = 0
  - Entidades nombradas
  - Expresiones numéricas
  - Palabras "raras"

intransigencia
e-book
a través de (expresion)

Análisis de la frecuencia

FI	Original	Simple
Palabras raras	9.49%	4.19%
NE	7.08%	8.77%
NumExp	2.81%	2.02%
Freq. 0 total	19.38%	14.98%
Freq. 3	1.23%	0.66%
Freq. 4	1.21%	0.89%
Freq. 5	6.02%	5.06%
Freq. 6	72.16%	78.40%

- Palabras con alto FI mas frecuentes en textos simples, palabras con bajo FI mas frecuentes en textos originales
- NE mas frecuentes en textos simples
- Palabras raras infrecuentes en textos simples
- Es probable que una combinación de frecuencia y longitud pueda usarse a la hora de seleccionar un reemplazo

Entidades nombradas y expresiones numéricas

- Varias operaciones con entidades nombradas y expresiones numéricas

Frases nominales definidas substituidas por sus referentes

la ciudad andaluza
↓
Granada

Redondear números grandes, eliminar números entre paréntesis, usar modificadores ...

El secretario general de la ONU, Ban Ki-moon, reclamó una mayor financiación mundial para atender las necesidades humanitarias de 2011, con un llamamiento de casi 7.400 millones de euros (unos 5.400 millones de euros), la mayor petición realizada jamás por Naciones Unidas.

↓

El secretario general de la ONU ha pedido más de 7000 millones de euros para las necesidades humanitarias.

Entidades nombradas y expresiones numéricas

- La operación mas común con expresiones numéricas es su eliminación (60%)
- Promedio de entidades nombradas mas grande en textos simples pues usualmente se repiten para que el mensaje sea mas claro
- Sin embargo si nos fijamos en entidades distintas, hay menos en los textos simples que en los complejos

Eliminación de oraciones

- Hemos identificado que no solo frases dentro de las oraciones se eliminan sino que oraciones completas no se incluyen en la simplificación
  - 20% de todas las oraciones son eliminadas
  - 72% de los textos tienen eliminación de oraciones (al menos una oración)
- La eliminación de oraciones puede tratarse como un problema de clasificación estadística
  - ✓ posición de la oración en el texto
  - ✓ # de palabras de contenido
  - ✓ # de signos de puntuación
  - ✓ # de nombres propios y expresiones numéricas
  - ✓ promedio de frecuencia de las palabras en la oración
  - ✓ conectividad entre frases vecinas

Eliminación de oraciones

- Se entrena un sistema utilizando el corpus Simplext y los rasgos identificados anteriormente
- Se incluyen dos "baselines":
  - (i) eliminar la última oración;
  - (ii) eliminar las dos últimas oraciones

Condition	Delete			Keep			Overall
	Prec.	Rec.	F1	Prec.	Rec.	F1	F1
Del. last	0.27	0.20	0.23	0.81	0.86	0.84	0.73
Del. 2 last	0.31	0.46	0.37	0.84	0.74	0.79	0.68
Classifier	0.42	0.26	0.30	0.86	0.89	0.87	0.79

Observaciones puntuales

- Substituciones
  - Todos los casos de verbos de "decir" ("afirmar", "explicar", "advertir", ...) son substituidos por *decir*.
  - Los adjetivos de nacionalidad (muy usuales en artículos de temática internacional) usualmente se substituyen por una forma perifrástica (pakistaníes -> personas de Pakistán)
  - Operaciones en expresiones numéricas:
    - Reemplazar una palabra por un número (cinco días -> 5 días)
    - Redondear números "grandes" (540.000 personas -> medio millón de personas)
    - "miles" y "millones" en números expresados en palabras (17.000 casas -> 17 mil casas)
  - Entidades nombradas
    - 90% de los nombres de persona en textos simples tienen tanto el nombre como el apellido(s)
    - 15% de los nombres de persona aparecen expandidos usando un calificativo "el pintor Pablo Picasso"

Otras simplificaciones

- Inserción de definiciones
  - Hemos observado que esta es una operación muy frecuente y productiva
  - Un poco mas de 57% de los textos simplificados contienen una definición que no aparece en el texto original
  - 70% de las definiciones corresponden a personas u organizaciones
  - un análisis de las definiciones encontradas en textos simples y definiciones extraídas de fuentes no simplificadas indica que:
    - las definiciones en textos simples usan palabras mas sencillas en términos de longitud
    - las definiciones en textos simples tienden a tener palabras con FI alto
    - en general no introducen nuevas entidades nombradas

Simplificación sintáctica

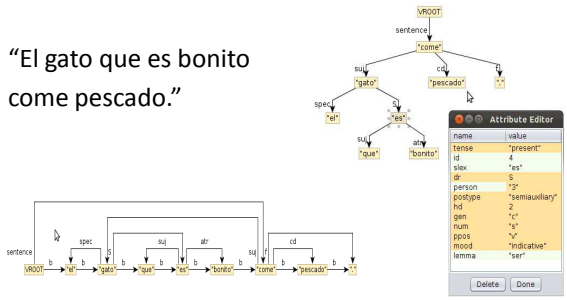
- Clausulas relativas simples y complejas
  - Las lluvias torrenciales, que comenzaron el pasado 1 de octubre y continuaron durante varios días, hicieron que los ríos y las presas se desbordaran, causando la peor inundación registrada en Vietnam en los últimos 20 años. -> *Las lluvias torrenciales causaron la peor… Estas lluvias comenzaron el 1 de octubre….*
- Construcciones participiales y de gerundio
  - Amnistía Internacional denunció que más de 150 personas del grupo opositor Hermandad Musulmana han sido detenidos este mes y antes de las elecciones parlamentarias, previstas para el 29 de noviembre. -> *Amnistía Internacional denunció que … antes de las elecciones parlamentarias. Estas elecciones están previstas para el 29 de noviembre.*
- Diferentes tipos de coordinación
  - Por este motivo, las personas no podían vivir en sus casas y fueron recogidas en campamentos. -> *Por este motivo las personas no podían vivir en sus casas. Estas personas fueron recogidas en campamentos.*

Herramientas para simplificar automáticamente

- Decidimos implementar un sistema que combina reglas con un filtro estadístico
  - La simplificación sintáctica de basa en un método de transformación de árboles de dependencias sintácticas
- Adoptamos un parser de dependencias sintácticas p/español (Bohnet, 2010) y la herramienta MATE para transducción de grafos
- El filtro estadístico es un componente basado en "Support Vector Machines" distribuidos con el sistema GATE

El parser

"El gato que es bonito come pescado."

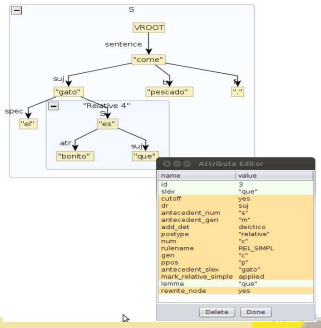


Reglas de transducción

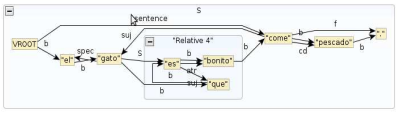
INPUT	OUTPUT	CONSTRAINTS

leftside = [
?Xl {
S-> ?Yl {
?r-> ?Zl
}
}
]
	rightside = [
	rc:?Zr {
	<=> ?Zl
	mark_relative_simple=applied
	}
	]
		conditions = [
		?Yl.ppos="v";
		?Yl.mood="indicative"
		?Zl.slex=que or ?Zl.slex=quién
		or ?Zl.slex=cuál
		or Zl.slex=donde;
		?Zl.id<?Yl.id;
		]

Aplicación de reglas



Transformación



Evaluación de las reglas

Simplification	Precision	Recall
clausulas relativas	0.40	0.80
de gerundio	0.63	0.20
coordinación de objetos	0.42	0.58
coordinación de verbos	0.64	0.50

- La evaluación se realiza sobre un conjunto de 886 oraciones del corpus
- Fuentes de error: parser erróneo, identificación errónea de clausulas, expresiones idiomáticas, etc.
- Decidimos estudiar la aplicación del filtro estadístico

Filtro estadístico

- Metodología
  - Recolectar oraciones de algún periódico (El País)
  - Parsear las oraciones y aplicar las reglas de simplificación para identificar donde deben ser aplicadas
  - Anotar las posiciones identificadas como correctas/incorrectas
  - Diseñar un clasificador para aprender cuando la decisión de simplificar es correcta

Filtro estadístico

- Evaluación por validación cruzada

Operation	Precision	Recall	F-Score
Relativas simples (mismo dominio)	0.85	0.86	0.86
Relativas simples (otro dominio)	0.76	0.76	0.76
Relativas complejas (mismo dominio)	0.70	0.71	0.71
Relativas complejas (otro dominio)	0.90	0.85	0.88

- Se comprobó el efecto del filtro estadístico en la aplicación de reglas de simplificación:
  - Mejora la efectividad en las oraciones de relativas complejas pero empeora la efectividad en las relativas simples

Evaluación

- Como la simplificación puede ser considerada como una tarea de generación de textos podremos evaluar el resultado contra la simplificación manual
- Como importa la legibilidad del resultado queremos tambien:
  - aplicar medidas de legibilidad a los textos y verificar si la complejidad de ha reducido
  - optaremos por aplicar alguna de las medidas que ya existen (Flesch, 1979); (Spauling, 1951)
- Tambien estamos evaluando (como hemos mostrando) la buena aplicación de reglas den estas o no un resultado idéntico al producido por un humano

Evaluación c/usuarios

- Evaluación extrínseca donde verificaremos tres condiciones textuales: textos originales, simplificaciones manuales, simplificaciones automáticas
  - Aplicación de tests de comprensión
  - Subjetos: individuos con sindrome de down con diferentes niveles de severidad
  - Contexto: programa PROMENTOR para promovel la inclusion de personas con discapacidad cognitiva en el mercado laboral y la sociedad – PRODIS foundation
  - Methodo: presentación de preguntas luego de la lectura de una version del texto

simplext Ejemplo de evaluación

- Texto original

Alex de la Iglesia dimitirá como presidente de la Academia de Cine
El director Alex de la Iglesia presentará su dimisión como presidente de la Academia de Cine tras la próxima gala de los Goya del próximo 13 de febrero.
Lo anuncia en un artículo publicado en la edición digital de "El País", en el que no vincula claramente su salida con su desacuerdo respecto a la "Ley Sinde", algo que sí ha sugerido en su página en Twitter.
Explica que ha hablado últimamente con representantes de sectores contrarios a su posición sobre las descargas en la Red y ha dialogado con ellos con fluidez.
"En este país, cambiar de opinión es el mayor de los pecados.
Creo que tenemos instalado el chip de la intransigencia desde hace tiempo", asegura.
"Lo coherente es dejarlo", seña el cineasta, que promete que seguirá defendiendo sus posiciones simplemente como director y que no "empañará" la gala de los Goya con su posición sobre la polémica de la Red.

simplext Ejemplo de evaluación

- Texto simplificado manualmente

Alex de la Iglesia dimitirá como presidente de la Academia de Cine
El director de cine Alex de la Iglesia deja la presidencia de la Academia de Cine.
Alex de la Iglesia dejará la presidencia de la Academia de Cine después de los Goya.
Los Goya es una fiesta donde se dan premios a los mejores actores y películas del cine español.
Alex de la Iglesia anunció su decisión en un artículo del periódico El País.
En su página web ha explicado que deja la Academia de Cine por no estar de acuerdo con la Ley Sinde.
La Ley Sinde es una ley que prohíbe las descargas gratis de películas en internet.
Alex de la Iglesia dimitirá como presidente de la Academia de Cine.

Ejemplo de evaluación

- Texto simplificado automáticamente

Alex de la Iglesia dimitirá como presidente de la Academia de Cine.
El director Alex de la Iglesia presentará su dimisión como presidente de la Academia de Cine tras la próxima gala de los Goya del próximo 13 de febrero.
Lo anuncia en un artículo publicado en la edición digital de "El País".
En este artículo no vincula claramente su salida con su desacuerdo respecto a la "Ley Sinde", algo que sí ha sugerido en su página en Twitter.
Explica que ha hablado últimamente con representantes de sectores contrarios a su posición sobre las descargas en la Red.
Ha dialogado con ellos con fluidez.
"En este país, cambiar de opinión es el mayor de los pecados."
Asegura: "Creo que tenemos instalado el chip de la intransigencia desde hace tiempo".
El cineasta seña: "Lo coherente es dejarlo".
Este cineasta promete que seguirá defendiendo sus posiciones simplemente como director y que no "empañará" la gala de los Goya con su posición sobre la polémica de la Red.

Ejemplo de evaluación

- Evaluación de la comprensión por preguntas/respuestas

1. ¿Qué cuenta la noticia que va hacer Alex de la Iglesia? (Dimitir / dejar de ser presidente de la Academia de Cine)
2. ¿Cuándo se retirará Alex de la Iglesia como presidente de la Academia de Cine? (Después los premios Goya, el 13 de Febrero)
3. ¿Dónde anunció Alex de la Iglesia que deja de ser presidente de la Academia de Cine? (En el periódico el País, en su página web de Twitter)
4. ¿Por qué se retira Alex de la Iglesia, como presidente de la Academia de cine? (Por no estar de acuerdo con la Ley Sinde)

Resumen

- Objetivo de la simplificación automática de textos
  - Transformar un texto en una versión que sea mas fácil de entender (y que tenga mas o menos el mismo contenido!)
- Hay un interés creciente en esta área en las comunidades de PLN, Recuperación de información, Accesibilidad Web.
- Dado que es una necesidad práctica y que no existe una solución para todos, existen muchos proyectos en desarrollo
- En los años a venir veremos aplicaciones que facilitaran el acceso a los textos en diversas lenguas

Resumen

- La simplificación automática de textos tiene por lo menos 3 ejes
  - Detección de la complejidad de un texto
  - Simplificación del vocabulario
  - Simplificación de la estructura
  - Varia técnicas se pueden aplicar: la tendencia en PLN es por métodos no supervisados o supervisados (cuando hay conjuntos de datos disponibles)
- La simplificación no es la única forma de hacer un texto mas accesible
  - Imágenes, tablas, figuras, animaciones
  - Disposición de los elementos textuales en los dispositivos
  - Habla

Ejercicios

- Calcular la complejidad de un texto en inglés usando la formula de Flesch ordenar los textos por dificultad (cuidado con las sílabas en inglés!)
  - Texto 1: Spain is a country in Southern Europe. It is in the Iberian Peninsula near Portugal and Gibraltar. France and the little country of Andorra are on its northeast side, where the Pyrenees mountains are.
  - Texto 2: **Spain**, officially the **Kingdom of Spain**, is a sovereign state and a member of the European Union located in southwestern Europe on the Iberian Peninsula. Its mainland is bordered to the south and east by the Mediterranean Sea except for a small land boundary with the British Overseas Territory of Gibraltar; to the north and north east by France, Andorra, and the Bay of Biscay; and to the northwest and west by the Atlantic Ocean and Portugal.

Ejercicios

- Simplificar el léxico de las siguientes oraciones
  - *"Juan regresó al hogar y se dio un remojón en la bañera."*
  - *"El leño ardían en el hogar de la casa de Juan"*
- Para la simplificación de las palabras podéis usar:
  - google usando búsquedas limitadas al dominio wikipedia en español (site:es.wikipedia.org)
  - WordReference para encontrar sinónimos de las palabras
- ¿Que problemas encontráis?

Ejercicios

- Simplificar la sintaxis de las siguientes oraciones
  - *"Juan regresó al hogar y se dio un remojón en la bañera."*
  - *"La niña que comió el pescado que fue pescado por la mañana se atragantó con una espina"*
- ¿Que problemas encontráis?

Ejercicios

- Simplificar lo mas posible (a nivel sintáctico y a nivel léxico) esta oración del corpus Simplex, si considera que hay algún elemento que se puede eliminar puede hacerlo
  - *"La escritora y académica Ana María Matute obtuvo el Premio Cervantes 2010, el mayor galardón de las letras españolas, que concede el Ministerio de Cultura con una dotación de 125.000 euros. "*
- ¿Que problemas encontráis?
- Como contribución a la ciencia os agradecería me enviarais vuestras simplificaciones a
  - horacio.saggion@upf.edu
  - Asunto: Simplificación Manual Curso Verano

Bibliografía

- 8 Sidor (sueco): www.8sidor.se
- Andrea Petz, Bror Tronback (2008) People with Specific Learning Difficulties: Easy to Read and HCI. ICCHP 2008: 690-692
- Barthe, Kathy; Juaneda, Claire; Leseigneur, Dominique; Loquet, Jean-Claude; Morin, Claude; Escande, Jean; Vayrette, Annick. (1999) GIFAS Rationalized French: A Controlled Language for Aerospace Documentation in French. Technical Communication, Volume 46, Number 2, May 1999 , pp. 220-229(10)
- Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, and Sandra Williams (2011). How to Make Numerical Information Accessible:
- Experimental Identification of Simplification Strategies. In INTERACT 2011, Part I, LNCS 6946, pp. 57–64, 2011.
- Bott S, Saggion H, Mille S. (2012). Text Simplification Tools for Spanish. Proceedings of Language Resources and Evaluation Conference, 2012.
- Bott S, Saggion H. (2011). An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction. ACL Workshop on Monolingual Text-to-Text Generation.
- Bott S, Saggion H. (2012). Automatic Simplification of Spanish Text for e-Accessibility. Proceedings of the 13th International Conference on Computers Helping People with Special Needs. :54–56.

Bibliografía

- Bouayad-Agha, N., Gerard Casamayor, Gabriela Ferraro, and Leo Wanner. (2009). Simplification of patent claim sentences for their paraphrasing and summarization. 2009 FLAIRS Conference.
- Carroll, J. Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. (1998). Practical Simplification of
- Chandrasekar, R. Christine Doran, and Bangalore Srinivas. 1996. Motivations and methods for text simplification. COLING 1996 , pps1041--1044.
- Coster, W. and David Kauchak. 2011. Learning to simplify sentences using wikipedia. Proceedings of Text-To-Text Generation, Portland, Oregon. Association for Computational Linguistics.
- De Belder, Jan; Deschacht, Koen; Moens, Marie-Francine. (2010) Lexical simplification, 1st international conference on interdisciplinary research on technology, education and communication, Kortrijk, Belgium, 25-27 May 2010
- Drndarevic B, Saggion H. (2012a). Towards Automatic Lexical Simplification in Spanish: An Empirical Study. Workshop Predicting and improving text readability for target reader populations (PITR2012). NAACL 2012.

Bibliografía

- Drndarevic B, Saggion H. (2012b). Drndarevic B, Saggion H. 2012. Reducing Text Complexity through Automatic Lexical Simplification: an Empirical Study for Spanish. SEPLN Journal . 49
- Easy-to-Read Network: http://www.easytoread-network.org/
- Enciclopedia Wikipedia (inglés): en.wikipedia.org
- Enciclopedia Wikipedia Simple (inglés): simple.wikipedia.org
- English Newspaper Text to Assist Aphasic Readers. AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology, pps7--10.
- Fellbaum, C. *WordNet*. An Electronic Lexical Database. The MIT Press.
- Inclusion Europe: http://inclusion-europe.org/
- Lectura Facil: http://www.lecturafacil.net
- Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, Lillian Lee (2010) For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. HLT-NAACL 2010: 365-368
- Mitkov, R. (2012) Natural Language Processing and Language Disabilities. Invited Talk. NLP4ITA LREC Workshop, Istambul. http://www.taln.upf.edu/nlp4ita/
- Noticias Fáciles (español): www.noticiasfacil.es
- Ogden, C.K. (1930) Basic English: A General Introduction with Rules and Grammar.

Bibliografía

- Or Biran, Samuel Brody, and Noémie Elhadad. Putting it Simply: a Context-Aware Approach to Lexical Simplification. 2011. ACL, pp. 496-501. Portland, OR.
- P. Lal and S. Rüger (2002) Extract-based Summarization and Simplification. In Proc. Workshop on Text Summarization Document Understanding Conference (DUC) 2002
- Luz Rello, Ricardo Baeza-Yates, Horacio Saggion, Jennifer Pedler (2012). A First Approach to the Creation of a Spanish Corpus of Dyslexic Texts. Proceedings of NLP4ITA. Istambul.
- Saggion H, Gómez-Martínez E, Etayo E, Anula A, Bourg L. (2011). Text Simplification in Simplext: Making Text More Accessible. Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural.
- Sandra M. Aluísio, Lucia Specia, Thiago Alexandre Salgueiro Pardo, Erick Galani Maziero, Renata Pontin de Mattos Fortes: Towards Brazilian Portuguese automatic text simplification systems. ACM Symposium on Document Engineering 2008: 240-248
- Siddharthan, A. (2002). An architecture for a text simplification system. Proceedings of LREC 2002. pp64-71

Bibliografía

- Spauling, S. 1956. "A Spanish Readability Formula" Modern Language Journal 40(8) 433-441.
- Vari-Cartier, P.: (1981): Development and Validation of a New Instrument to Asses the Readability of Spanish Prose, Modern Language Journal, 65, 141-148.
- Walter Daelemans and Anja H¨othker and Erik Tjong Kim Sang. (2004). Automatic Sentence Simplification for Subtitling in Dutch and English. Proceedings of the Language Resources and Evaluation Conference.
- Zhu, Z., Delphine Bernhard, and Iryna Gurevych. (2010). A monolingual tree-based translation model for sentence simplification. Proceedings of The 23rd International Conference on Computational Linguistics, pps1353--1361, Beijing, China, Aug.