



Erros Linguísticos e Qualidade de Sumários

Lucia Castro

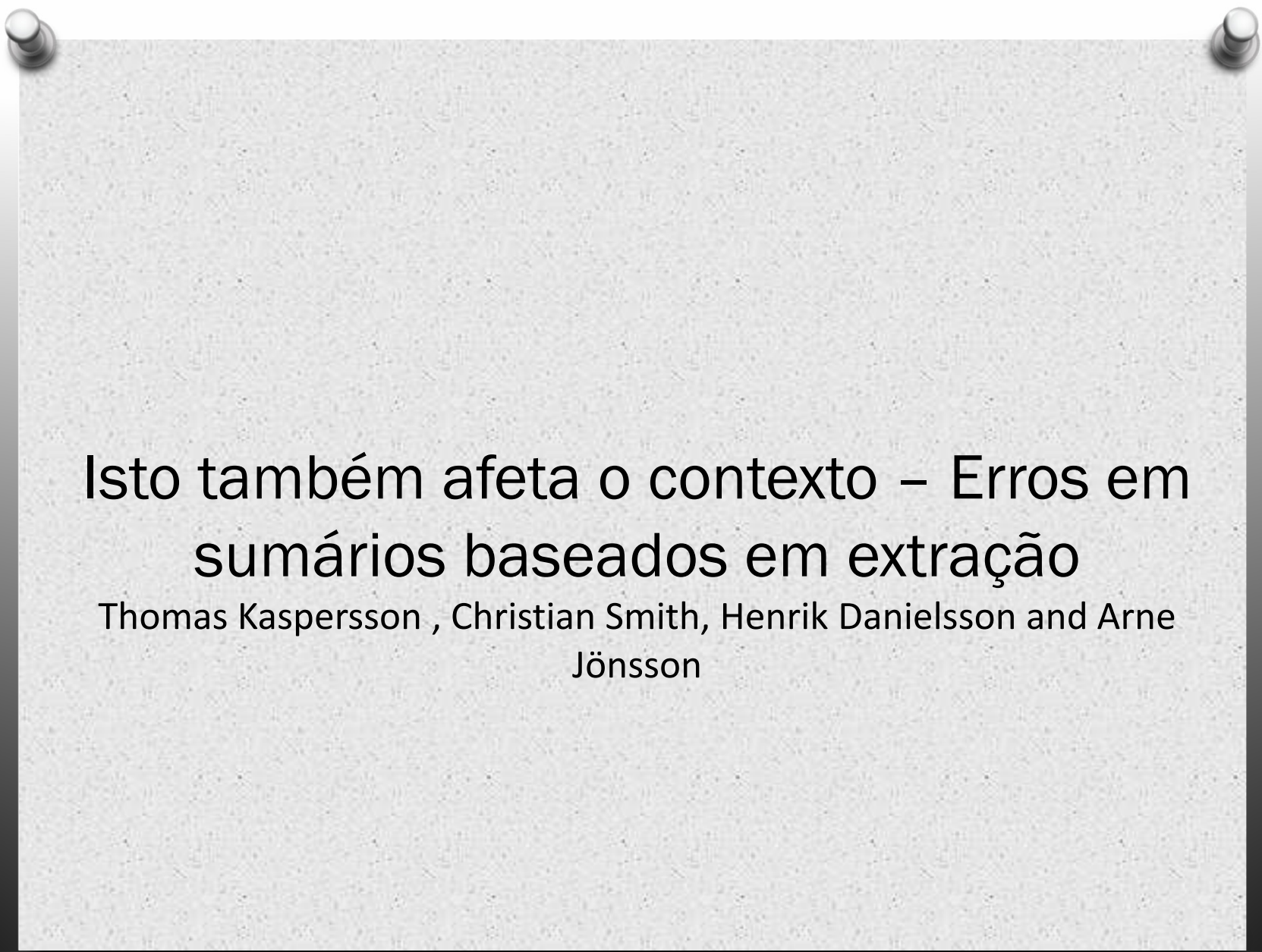
Márcio Dias

Introdução

- o Um sumário baseado em extração extrai a sentença mais importante;
- o Tal extração pode provocar diversos erros linguísticos que afetam a qualidade do sumário

Introdução

- o Serão apresentados dois estudos sobre erros linguísticos e qualidade de sumários
- o Isto também afeta o contexto – Erros em sumários baseados em extração. Kasperson et al. (2012)
- o Avaliação Automática da Qualidade Linguística em Sumários Multidocumento. Pitler et al. (2012)



Isto também afeta o contexto – Erros em sumários baseados em extração

Thomas Kaspersson , Christian Smith, Henrik Danielsson and Arne Jönsson

Motivação

- o Poucos estudos investigam a distribuição de erros no sumário
- o Impacto dos erros com diferentes tamanhos de sumário (níveis de sumário)
 - o Será que tem influência?

Contribuição

- o O estudo apresenta uma investigação dos erros linguísticos que ocorrem em sumários monodocumentos.

Metodologia

- o Os erros foram encontrados por meio da seguinte tarefa:
 - o Ler os textos sumarizados de **3 gêneros** e com **5 níveis** (tamanhos) diferentes;
 - o Os Gêneros:
 - o Textos Jornalísticos (190 palavras por texto);
 - o Textos de Científicos (650 palavras por artigo)
 - o Textos Oficiais (Autoridade – 720 palavras por texto)
 - o Os níveis de sumarização
 - o 17%, 33%, 50%, 67% e 83%
 - o Marcar nos textos os erros encontrados com a sua respectiva descrição

Metodologia

- o Marcação foi feita por analisadores e para cada um foi dado quatro textos sumarizados;
- o Os **erros** encontrados foram agrupados em **3 categorias** e subcategorias:
 - 1) Referência anafórica errônea
 - (a) Sintagma Nominal
 - (b) Nomes Próprios
 - (c) Pronomes
 - 2) Ausência de coesão ou contexto
 - 3) Referência anafórica quebrada
 - (a) Sintagma Nominal
 - (b) Nomes Próprios
 - (c) Pronomes


Erros em sumários baseados em extração

o Exemplo 1

Originally the return from Uppsala royal estate property should be enough for the kings support. What we nowadays call taxes was not in question - the free man could not be forced to pay any fees. The free man had, however, official duty.

Such official duty was the guesting, the obligation to receive and support the king and his escort when they travelled.

o Exemplo 2



De Long and his crew sailed with the ship Jeannette through the Bearing sea 1879.

He let build a powerful ship strong enough to drift unharmed with the thick pack ice for a long time.

o Exemplo 2

But the trip towards the north pole became a disaster.

De Long and his crew sailed with the ship Jeannette through the Bearing sea 1879.

Soon they got stuck in ice north of the Wrangel island. In June 1881, Jeannette was crushed by the ice, and everyone onboard perished after a time of hardship. The theory about the open polar sea was declared dead. The disaster however, became of great importance for polar research. A few years after the foundering of the Jeannette wreck parts reached the east coast of Greenland - a revolutionary discovery. Fridtjof Nansen immediately got the idea to test the theory of an open sea filled with drift ice

He let build a powerful ship strong enough to drift unharmed with the thick pack ice for a long time.

Carried by the ice, the expedition would travel from Siberia to the North pole.

Erros em sumários baseados em extração

o Exemplo do CSTNews

De acordo com a pesquisa, Lula (PT) tem 44% das intenções de voto, contra 25% de Geraldo Alckmin (PSDB) e 11% de Heloísa Helena (PSOL). Heloísa Helena, candidata à Presidência pelo PSOL, aparece em terceiro, com 11% das intenções de voto, seguida por Cristovam Buarque (PDT) e Luciano Bivar (PSL), ambos com 1%.

O Instituto de pesquisa CNI/Ibope divulgou nesta sexta-feira, em Brasília, que o presidente Luiz Inácio Lula da Silva seria reeleito em primeiro turno caso as eleições fossem hoje.

A pesquisa CNI/Ibope realizada em julho e divulgada nesta sexta-feira mostra que o presidente Luiz Inácio da Silva teria 44% dos votos no primeiro turno, enquanto o candidato tucano Geraldo Alckmin teria 25% das intenções de voto.

■ Erro de referência anafórica quebrada envolvendo os D2_C2 e D4_C2

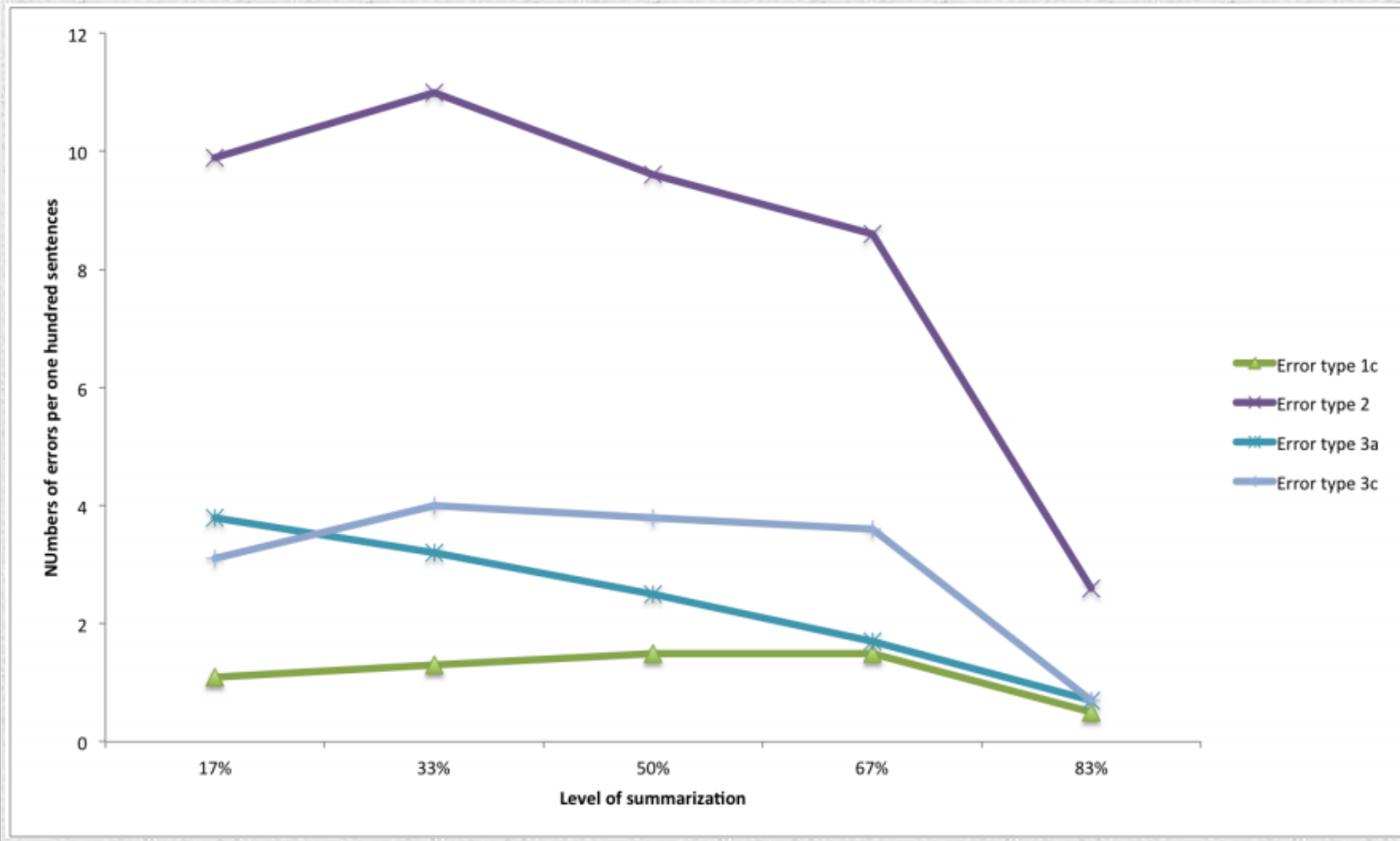
Metodologia de Avaliação

- o 30 textos (10 de cada gênero) e os 5 diferentes tamanhos do sumário;
- o Textos são apresentados linha por linha, com sentenças extraídas em negrito e sentenças não extraídas marcadas em vermelho;
- o 2 pares de leitores foram utilizados para marcar os erros nos 30 textos;
- o Foram dados 20 textos para cada par, ou seja, 10 textos dos 30 eram os mesmos, sendo obtido 69,4% de grau de confiança.

Resultados

- Não foram encontrados diferenças significativas na quantidade de erros entre os gêneros;
- Os tipos de erros mais significativos foram:
 - Referência anafórica errônea, subtipo pronome (1c)
 - Ausência de coesão e contexto (2)
 - Referência anafórica quebrada, sintagma nominal (3a)
 - Referência anafórica quebrada, pronome (3c)

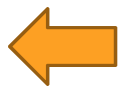
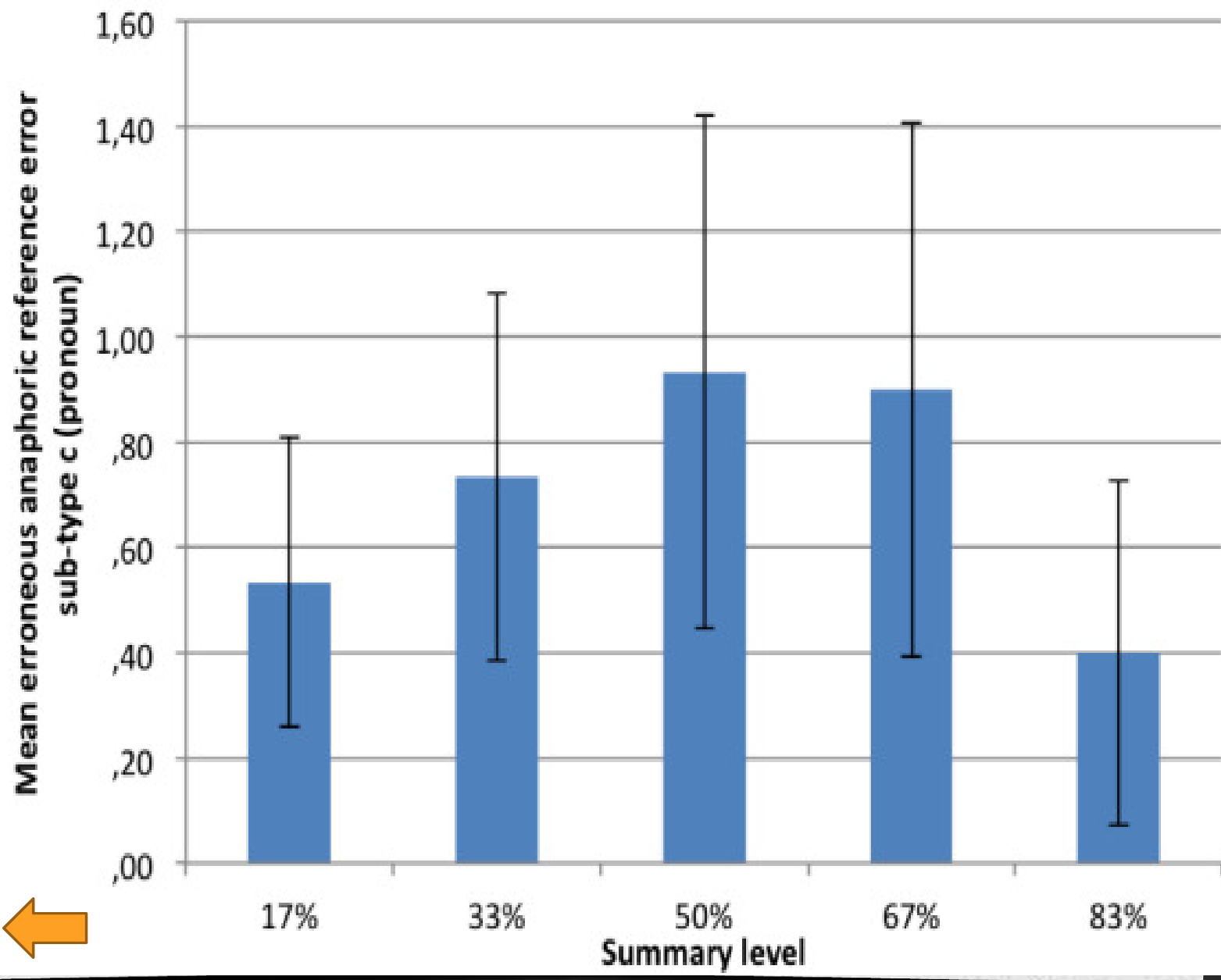
Discussões

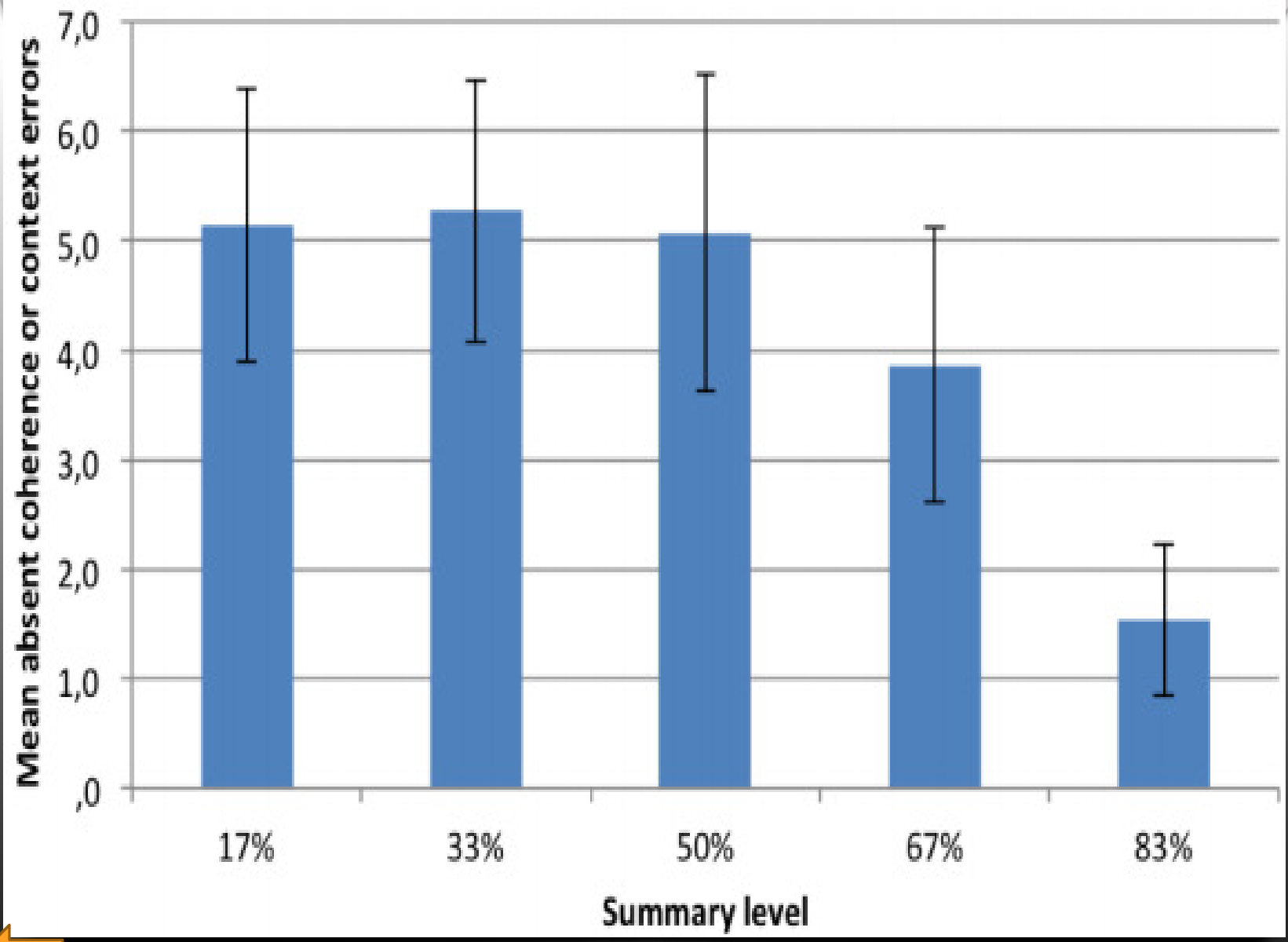


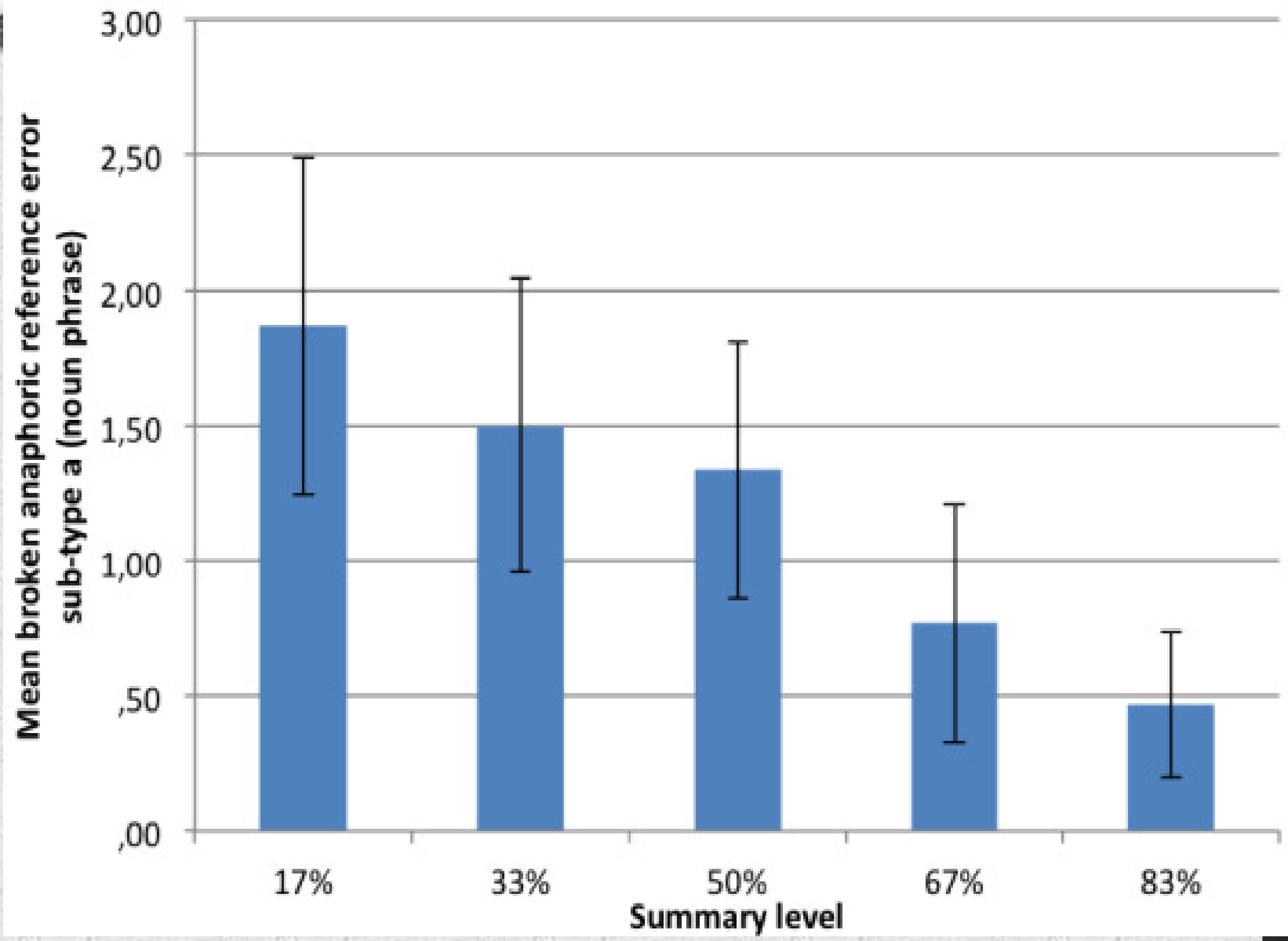
- o Os resultados mostram que os erros mais comuns são ausência de coesão ou contexto e vários tipos de falta de referência anafórica e de quebra;
- o Os autores acreditam que resultados obtidos são relevantes para qualquer sumário baseado em extração
- o Diferentes tipos de erros, mostram relações bem diferentes para os tamanhos de sumário considerados.

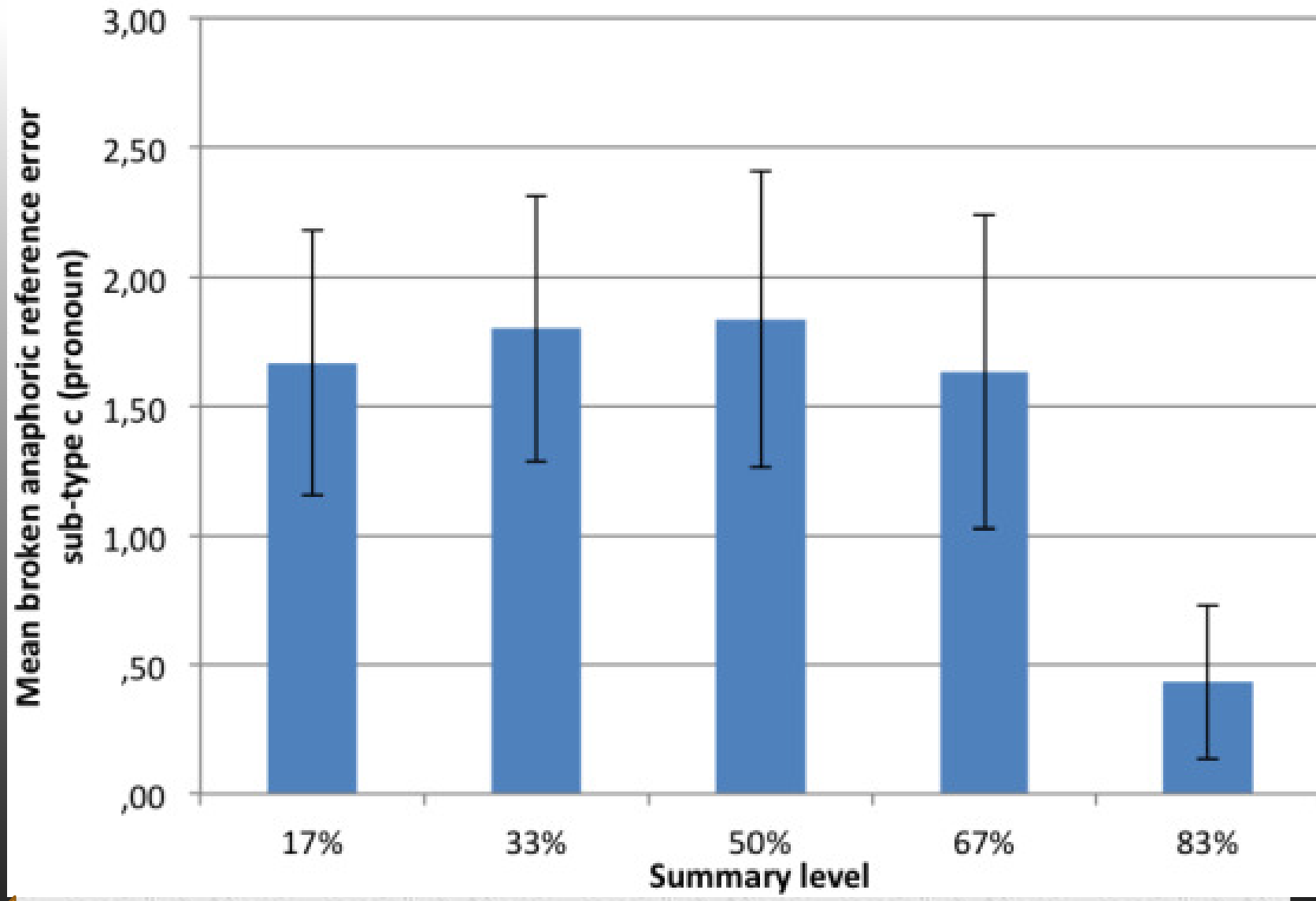
- Os resultados mostram que tamanho do sumário tem que ser levado em consideração para minimizar a quantidade de erros produzidos pelo sumarizador extrativo.
- Os tipos de erros vistos afetam a coerência e o discurso e frequentemente faz o texto difícil de ler ou incompreensivo.
- Este estudo também elucidada a importância de melhorar a geração textual para sumarizadores extrativos.

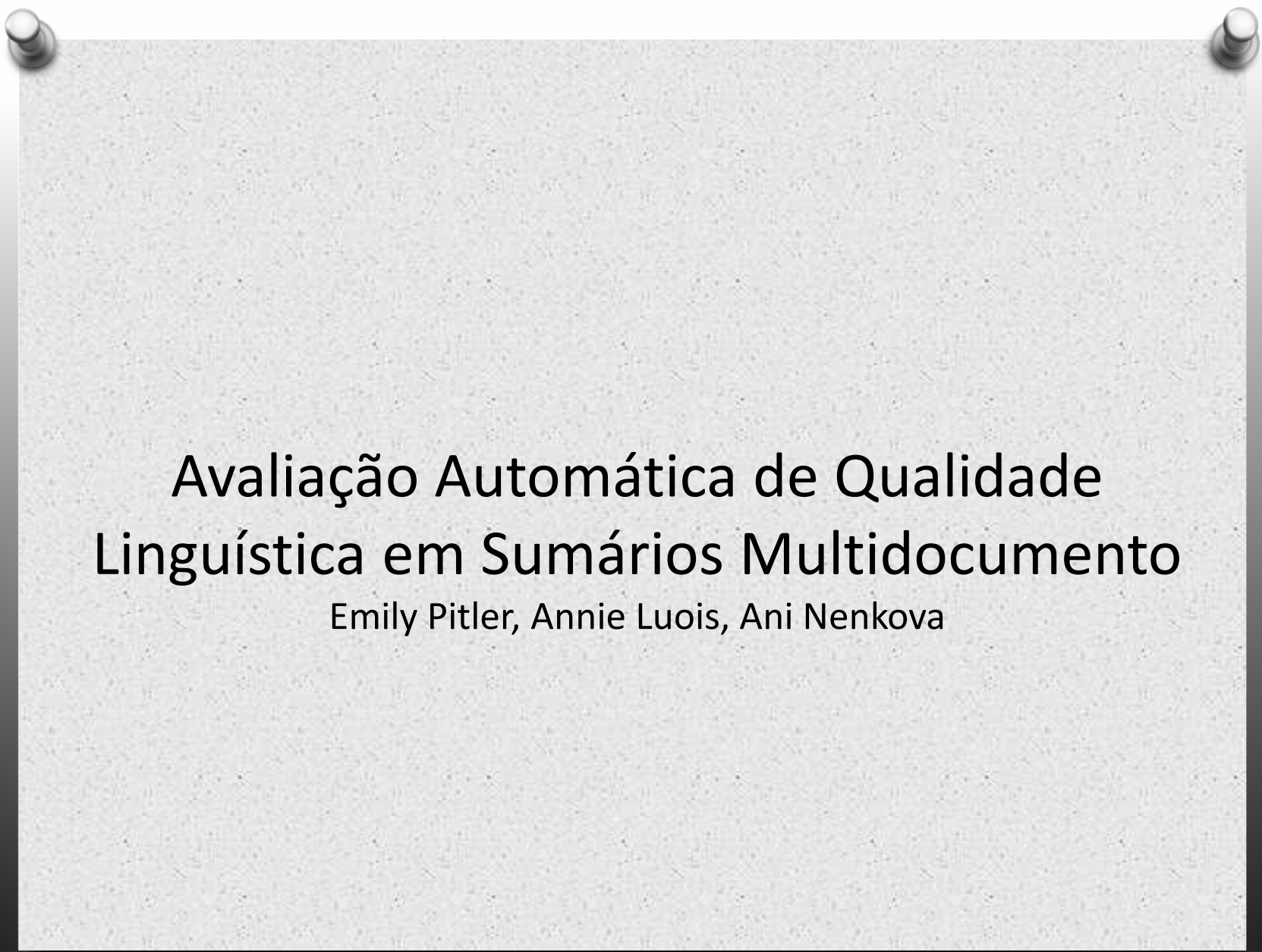












Avaliação Automática de Qualidade Linguística em Sumários Multidocumento

Emily Pitler, Annie Luo, Ani Nenkova

Objetivos do Trabalho

- Estudo da influência de vários tipos de características na avaliação automática de sumários multidocumento
- Treinamento de modelos
- Ranque de sumários com base na qualidade

Aspectos da qualidade linguística

- Aspectos da qualidade linguística de sumários definidos pela TAC (Text Analysis Conference)
 - *Gramaticalidade;*
 - *Não-Redundância;*
 - *Clareza Referencial;*
 - *Foco;*
 - *Estrutura e coerência.*

Indicadores de Qualidade linguística

- Modelos de língua
 - 1-grama, 2-grama 3-grama
 - c/ngrama → min, max e average log probability de cada sentença
- Entidades Nomeadas(EN)
 - Informações descritivas no sumário?
 - Primeiras menções de EM são mais descritivas
 - Features → nro de PM no sumário que não PM no texto fonte

Indicadores de Qualidade linguística

- **Sintaxe de Sintagmas Nominais**
 - Média de c/ POS e c/ frase sintática no sintagma
- **Formas de coesão**
 - Nro de demonstrativos, pronomes e conectores discursivos
- **Continuidade**
 - Expressões referenciais e conectivos discursivos
 - Contexto sumário e textos fonte → é igual?
 - Nro de vezes onde Sent. preced Sum = Sent. preced Texto Fonte; viceversa

Indicadores de Qualidade linguística

- **Sintaxe de Sintagmas Nominais**
 - Média de c/ POS e c/ frase sintática no sintagma
- **Formas de coesão**
 - # demonstrativos, pronomes e conectores discursivos
- **Continuidade**
 - Expressões referenciais e conectivos discursivos
 - Contexto sumário e textos fonte → é igual?
 - Nro de vezes onde Sent. preced Sum= Sent. preced Texto Fonte; viceversa
 - Medida de similaridade coseno entre sent.adj.
 - # vezes o antecedente do pronome está antes da ou na sentença atual

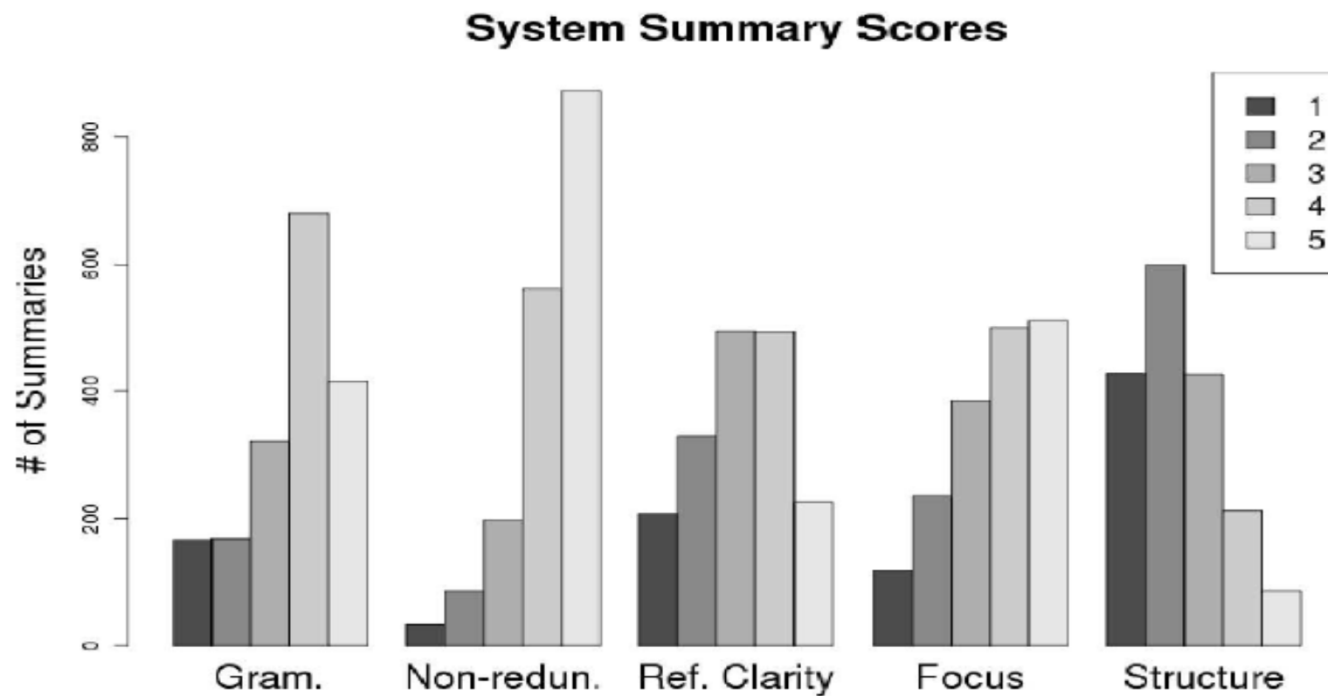
Indicadores de Qualidade linguística

- **Fluência Sentencial**
 - tam. da sentença, tam. médio das frases, # de frases
- **Coh-matrix (Graesser et al., 2004)**
 - #silabas de palavras, LSA sent. Adjacentes, etc.
- **Coerência de Palavras (Soricut e Marcu, 2006)**
 - Probabilidade de palavras em sentenças adyacentes
- **Coerência de entidades**
 - Modelo de Barzilay e Lapata (2008)→ *Entity Grid model*

Avaliação Manual da qualidade dos sumários

- Avaliação realizada sobre os dados da DUC 2006
 - 50 entradas
 - 35 sumarizadores
 - Sumários de 250 palavras
- C/ sumário foi avaliado com os 5 aspectos da qualidade linguística
 - Nota 1(ruim) até 5(muito bom)

Avaliação Manual da qualidade dos sumários



Correlação de Spearman sobre as pontuações manuais

	Gram	Non-redun	Ref	Focus	Struct
Content	.02	-.40 *	.29	.28	.09
Gram		.38 *	.25	.24	.54 *
Non-redun			-.07	-.09	.27
Ref				.89 *	.76 *
Focus					.80 *

Metodologia experimental

- Treinamento → DUC 2006
- Teste → Duc 2007
- Medida de avaliação → acurácia
 - Pares de sumários ranqueados corretamente no conjunto de teste
- Usou-se SVM-light que é um svm para ranqueamento
 - O objetivo desse ranqueador é minimizar o numero de pares discordantes com o ranqueamento ideal
 - A saída do ranqueador é um score (c/sentença)
 - **Aplica-se um meta ranqueador que combina os scores para todo o sumário**

Resultados

Feature set	Gram.	Redun.	Ref.	Focus	Struct.
Lang. models	87.6	83.0	91.2	85.2	86.3
Named ent.	78.5	83.6	82.1	74.0	69.6
NP syntax	85.0	83.8	87.0	76.6	79.2
Coh. devices	82.1	79.5	82.7	82.3	83.7
Continuity	88.8	88.5	92.9	89.2	91.4
Sent. fluency	91.7	78.9	87.6	82.3	84.9
Coh-Metrix	87.2	86.0	88.6	83.9	86.3
Word coh.	81.7	76.0	87.8	81.7	79.0
Entity coh.	90.2	88.1	89.6	85.0	87.1
Meta ranker	92.9	87.9	91.9	87.8	90.0

Table 2: System-level prediction accuracies (%)

Resultados

Feature set	Gram.	Redun.	Ref.	Focus	Struct.
Lang. models	66.3	57.6	62.2	60.5	62.5
Named ent.	52.9	54.4	60.0	54.1	52.5
NP Syntax	59.0	50.8	59.1	54.5	55.1
Coh. devices	56.8	54.4	55.2	52.7	53.6
Continuity	61.7	62.5	69.7	65.4	70.4
Sent. fluency	69.4	52.5	64.4	61.9	62.6
Coh-Metrix	65.5	67.6	67.9	63.0	62.4
Word coh.	54.7	55.5	53.3	53.2	53.7
Entity coh.	61.3	62.0	64.3	64.2	63.6
Meta ranker	71.0	68.6	73.1	67.4	70.7

Table 3: Input-level prediction accuracies (%)

Resultados

Feature set	Gram.	Redun.	Ref.	Focus	Struct.
Lang. models	52.1	60.8	76.5	71.9	78.4
Named ent.	62.5	66.7	47.1	43.9	59.1
NP Syntax	64.6	49.0	43.1	49.1	58.0
Coh. devices	54.2	68.6	66.7	49.1	64.8
Continuity	54.2	49.0	62.7	61.4	71.6
Sent. fluency	54.2	64.7	80.4	71.9	72.7
Coh-Matrix	54.2	52.9	68.6	56.1	69.3
Word coh.	62.5	58.8	62.7	70.2	60.2
Entity coh.	45.8	49.0	54.9	52.6	56.8
Meta ranker	62.5	56.9	80.4	50.9	67.0

Table 5: Input-level prediction accuracies for human-written summaries (%)

Conclusões

- Acurácias altas para uma entrada e para o teste completo do sistema confirmam que a avaliação da qualidade de sumários pode ser realizada satisfatoriamente usando técnicas computacionais

Referências Bibliográficas

(HASSEL,2000) HASSEL, M. *Pronominal resolution in automatic text summarisation*. Master's thesis, Master thesis in Computer Science, Department of Computer and Systems (DSV), Stockholm University, Sweden, 2000.

(Kaspersson et al., 2012) Kaspersson, T.; Smith, C.; Danielson, H.; Jönsson, A. This also affects the context - Errors in extraction based summaries. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Turkey.

(OTTERBACHER,2002) OTTERBACHER, J. C.; RADEV, D. R. and LUO, A. *Revisions that improve cohesion in multidocument summaries: A preliminary study*. In Proceedings of the Workshop on Automatic Summarization (including DUC 2002), Philadelphia, pages 27-30, 2002.

(OVER,2007) OVER, P.; DANG, H. and HARMAN, D. *Duc in context*. *Information Processing & Management*, 43:1506-1520, Janeiro, 2007.

(PITLER ET AL., 2012) . PITLER, E. ;LOUIS, A.; NENKOVA, A. Avaliação Automática de Qualidade Linguística em Sumários Multidocumento

(SMITH and JONSSON,2011) SMITH, C. and JONSSON, A. *Automatic summarization as means of simplifying texts, na evaluation for swedish*. In Proceedings of the 18 th Nordic Conference of Computational Linguistics (NoDaLiDa-2010), Riga, Latvia, 2011.