INTRODUÇÃO

Justificativa

Este projeto se insere no contexto da sumarização automática de textos, tópico cujo interesse foi retomado nas últimas décadas com o advento da internet, pela necessidade de se reter grandes volumes de informações em curtíssimos espaços de tempo. A sumarização automática, nesse contexto, pode servir, por exemplo, para o usuário recuperar informações que lhe sejam relevantes.

Objetivos

Este projeto visa a prototipação de um sumarizador automático baseado na modelagem de discurso de Rino (1996), cuja arquitetura segue o modelo tradicional da geração de textos de dois passos: planejamento textual e realização lingüística, conforme mostra a Figura 1. O primeiro passo é responsável por gerar o(s) plano(s) de texto a partir da interpretação de um texto-fonte (que compõe a base de conhecimento do sistema), enquanto o segundo se responsabiliza por gerar a expressão lingüística do(s) plano(s) de texto por meio de escolhas morfossintáticas.

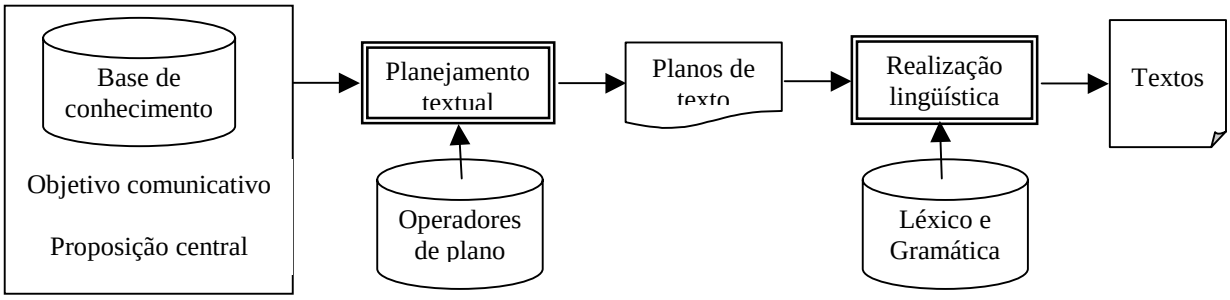

Figura 1 – Arquitetura do gerador de textos de dois passos

REVISÃO LITERÁRIA

Neste trabalho, o planejamento textual, baseado em operadores de plano, é dirigido pela satisfação do objetivo comunicativo do discurso. Este, por sua vez, depende da satisfação de outros subobjetivos. Esse planejamento é feito, portanto, incrementalmente, aplicando-se operadores de plano sucessivamente e gerando, ao mesmo tempo, o plano de texto. Os operadores de plano seguem o modelo sugerido por Moore e Paris (1993) com base na RST (*Rhetorical Structure Theory* – Mann and Thompson, 1987). O texto-fonte, por sua vez, consiste em uma estrutura puramente informativa, representada segundo o modelo Problema-Solução (Jordan, 1980; Winter, 1976). Para complementar o planejamento, a decisão sobre relações retóricas toma por base também as relações intencionais da GSDT (*Grosz and Sidner Discourse Theory* – Grosz and Sidner, 1986). Desse modo, o modelo de planejamento textual de Rino é incorporado a este trabalho.

METODOLOGIA

A metodologia de desenvolvimento do protótipo do sumarizador automático é baseada em exemplos, cujos textos-fonte dão origem à base de conhecimento que, juntamente com o objetivo comunicativo e a proposição central, formam a entrada do sistema.

A partir da especificação e variação dos componentes de entrada do sistema, com base no mapeamento entre as relações discursivas de Rino, foram definidos grupos de operadores de plano, que são artifícios computacionais para a determinação da estrutura e do conteúdo de um texto, sendo os mesmos o ponto de partida para a implementação do planejamento textual. Este, por sua vez, se dá através do processamento profundo do discurso, mapeando relações intencionais e semânticas em relações retóricas, buscando manter a coerência e a coesão do texto a ser gerado.

O planejamento se inicia pela escolha dos operadores de plano potenciais que satisfaçam o objetivo comunicativo, preservando a proposição central. Esse processo é guiado pela associação das relações de discurso nos três níveis de representação (semântica, intencional e retórica), a partir da definição tabelada por Rino, conforme exemplificado na Tabela 1.

Tabela 1 – Mapeamento de relações semânticas e intencionais em relações retóricas

Relações semânticas	**Relações retóricas**	**Relações intencionais**
enable(X,Y)	*means*(Y,X)	X *sat-precedes* Y Y *dom* X

Para o planejamento, essa tabela é usada da seguinte forma: caso se tenha uma relação *enable* entre dois segmentos de discurso *X* e *Y* na base de conhecimento e a verificação das relações intencionais *X sat-precedes Y* e *Y dom X*, produz-se a relação retórica *means* no plano de texto. A Figura 2 mostra um exemplo de operador definido para esse caso, que satisfaz, por exemplo, o objetivo comunicativo *descrever* com uma proposição central qualquer representada por X.

name: descrever-complexo
header: descrever(X)
effect: know-about(reader,X)
constraints: not know-about(reader,X), isa(X,Y,*enable*), X *sat-precedes* Y, Y *dom* X
nucleus: descrever(X)
satellite: know(reader,*means*(X,Y))

Figura 2 – Operador de plano para descrição de um conceito

RESULTADOS

Foram feitos cinco estudos de caso para a implementação do planejamento textual, resultando em um total de 52 operadores de plano. Com isso, foram produzidos cerca de 250 planos de texto que, em primeira análise, são coerentes. Tais planos, por sua vez, serão realizados lingüisticamente no próximo estágio deste projeto. A Figura 3 mostra uma interface do planejador e o plano de texto gerado pelo operador de plano acima (o segmento de discurso 1 corresponde a X, enquanto o 2 corresponde a Y) para o texto abaixo (Rino, 1996, p. 261), cujo sumário pode ser "*We studied the laser processing of materials by measuring the depth of the keyhole structure when a cw CO_2 laser beam was focused onto water.*" (Rino, 1996, p. 266).

1. *This paper presents the results of measurements on the keyhole structure produced during the laser processing of materials.*
2. *A cw CO_2 laser was used to produce keyhole structures in water and measurements were made of the depth of the keyhole as a function of laser power and the force due to gravity.*
3. *The data show that as the laser power is increased, the keyhole depth also increases.*
4. *However, changes in the force due to gravity seem to have little effect on the depth*
5. *The data for the different power conditions are compared with calculations using the Andrews and Atthey expression for the hole shape and good agreement is achieved.*

Figura 3 – Planejador textual

CONCLUSÕES E COMENTÁRIOS

Este trabalho é a implementação de um modelo que pretende cobrir os problemas de seleção e estruturação de informações de forma cega, isto é, problemas de coerência e coesão, a partir de um repositório informativo. Os textos gerados deverão, portanto, ser avaliados segundo esta perspectiva, verificando-se, principalmente, a preservação da proposição central e a satisfação do objetivo comunicativo, o que pode se dar de várias formas: comparando-se o plano de texto (janela direita da Figura 3) com a base de conhecimento (janela esquerda da Figura 3), comparando-se o sumário final com os sumários gerados por métodos empíricos (cegos), entre outras.

REFERÊNCIAS

Grosz, B. and Sidner, C. Attention, intentions, and the structure of discourse. In *Computational Linguistics*, Vol. 12, No. 3, 1986.

Jordan, M. P. Short texts to explain problem-solution structures – and vice versa. In *Instructional Science*, Vol. 9, pp. 221-252, 1980.

Mann, W.C. and Thompson, S.A. *Rhetorical structure theory: a theory of text organization*. Technical Report ISI/RS-87-190, 1987.

Moore, J.D. and Paris, C. Plannig text for advisory dialogues: capturing intentional and rhetorical information. In *Computational Linguistics*, Vol. 19, No. 4, pp. 651-694, 1993

Rino, L.H.M. *Modelagem de discurso para o tratamento da concisão e preservação da idéia central na geração de textos*. Tese de Doutorado. IFSC-Usp. São Carlos – SP, 1996.

Winter, E. O. *Fundamentals of information structure*. Hatfield Polytechnic, Hertfodshire, England, 1976.

FINANCIAMENTO

FAPESP – bolsa de mestrado, processo no. 99/11954-6.

AGRADECIMENTOS

Ao NILC – Núcleo Interinstitucional de Lingüística Computacional, formado por pesquisadores da UFSCar, USP/São Carlos e UNESP/Araraquara.