

## Não deitar fora o bebé com a água

Diana Santos  
Linguateca  
[www.linguateca.pt](http://www.linguateca.pt)

## Motivação; grandes linhas

- Transmitir algumas das minhas reflexões e do meu conhecimento
- Forçar à reflexão
- O que é probabilidade, frequência, distribuição
- Porque é que se conta?
- Introdução à linguística e ao PLN
  - O que é que a língua tem de especial?

## Linguística vs. PLN

- Linguística: Estudo da linguagem humana como uma ciência
- generalizações, previsões, refutações
  - uma teoria é tanto melhor quanto o seu potencial de refutabilidade
- com o intuito de melhorar
  - a interação entre humanos (retórica, ensino, tradução)
  - a interação mediada por computadores (processamento)
- PLN: a engenharia da linguística
- com base nas teorias, desenvolver sistemas que mudem / interajam com o mundo
  - diferença entre a ciência, que observa e explica, e a engenharia, que manipula e modifica

## A questão do quantitativo

- Primeira questão: o que é que se conta?
- palavras, frases, expressões, sílabas, letras, actos de fala, erros, ...
- é preciso um modelo / uma teoria para cada uma destas unidades
- é preciso um modelo de para que é que as contagens (frequências) servem
- não se devia estar a fazer contagens antes de explicar porquê
- Listas de frequências

## Porque é que se conta?

- Quais são as hipóteses “escondidas” para contar?
- para poder prever o futuro?
- para poder aplicar métodos estatísticos?
  - e quais são as bases para estes?
- em que é que uma frase dita só uma vez ou uma frase repetida mil vezes diferem?
- em que é que uma palavra muito frequente difere de uma pouco frequente?
- O que é que significam médias?
- O que é que significam probabilidades?

## Em resumo, em que é que os métodos quantitativos diferem dos qualitativos?

- Se tivermos regras que descrevem uma análise
- o que é que ajuda termos a frequência com que essas regras são aplicadas?
  - ajuda quando não soubermos
- São coisas diferentes
  - Métodos quantitativos no funcionamento do sistema
  - Métodos quantitativos na avaliação do sistema
  - Métodos quantitativos no estudo de um problema
- Chocolates do Carbonell

## Sintaxe vs. semântica

- Na língua há sempre duas faces, igualmente importantes
  - a sintaxe: como exprimir: ordem, concordância
  - a semântica: o quê exprimir, quais os participantes e o que se deixa implícito
  - exemplo: saber conduzir e ir a algum sítio
- Mas a propriedade mais essencial da língua é que tem uma **intenção**
- Um dos grandes problemas da linguística é ter separado demasiado estas duas (ou três) questões (dividir para reinar), ao contrário da IA
  - Chomsky: as ideias coloridas dormem verdemente
  - Hintikka: é preciso critérios semânticos para decidir a aceitabilidade sintáctica
  - Lakoff & Johnson: as próprias unidades da sintaxe requerem compreensão
  - *É preciso compreender para atribuir um valor de verdade*

## Lógica filosófica e língua

- Infelizmente, também a filosofia não ajudou muito, visto que a lógica se preocupou sobretudo com
  - valores de verdade
  - quantificação (ontologia)
  - modalidades (deónticas, doxásticas, etc.)
- embora inspirada pela linguagem natural!
- Os filósofos lógicos queriam um modelo do raciocínio ideal, não um modelo da forma como as pessoas comunicam
  - 1. existência não é um predicado, 2. tudo existe, 3. há coisas que não existem
  - Hirst propõe 9 tipos diferentes de existência para RC (representação do conhecimento)

## Formalismo vs. linguagem natural

- Todos os homens são estúpidos!
- $\forall x \text{ x in H} \Rightarrow \text{estupido}(x)$
- Diferenças
  - filho bebé
  - Papa
  - o homem ideal
  - D'Artagnan
  - o homem de 3050
  - André Gide
- A intenção da frase é o mais importante! depende do falante, do contexto, ...

## Diferenças entre a LN e as LAs

- Ou melhor, as LAs (línguas criadas conscientemente pelo homem) são uma forma de
  - simplificar a expressividade da LN
  - corrigir os problemas da LN
  - ignorar a complexidade da LN
- A riqueza da LN compreende:
  - vagueza
  - dependência do contexto
  - referência a conhecimento implícito
  - carácter dinâmico (evolução e criatividade)
  - carácter metafórico
  - a intenção de **comunicação** (diálogo)

## Metáfora: propriedade constitutiva da língua

- Mudança de paradigma proposta por Lakoff & Johnson (1980)
- Percebemos um domínio em termos de outro
- O sistema conceptual está organizado metaforicamente
- Metáforas culturais fazem parte da nossa forma de ver o mundo
  - Problema como puzzle
  - Problema como precipitado numa solução química
  - Argumentação como guerra
- É preciso compreender para atribuir um valor de verdade
  - *A teoria da relatividade estava mesmo à mão para Heisenberg*
- 3 tipos de metáforas convencionais: estruturais, orientacionais, e ontológicas

## Metáfora como inspiração em informática

- Metaphors we surf the Web by
- Desenho de interfaces em Macs

## Dependência do contexto

- O sentido das palavras/frases/enunciados depende **sempre** do contexto, em LN
- *o livro*
- deícticos, quase deícticos, quase referenciais, anafóricos, etc.
- *aquí, ali, lá (posição), eu, tu, nós (falante), hoje, agora (tempo)*
- *depois, antes*
- *outro, este, primeiro*
- *presidente da República, presidente da mesa, o país, a assembleia,*
- *pai, meu*

## Dependência do contexto em LPs

- $a=5$  e  $b=a$ ,  $a$  significa em geral coisas diferentes
- $a+3$  e  $\text{substr}(a)$  idem (coacção)
- $a="aaa"; \text{print } a$  e  $a="bbb"; \text{print } a$ , tb dá um resultado diferente
- assim como  $\text{print "print"}$  ou mesmo  $\text{print PRINT "print \$print"}$

## Dependência do contexto 2

- Exactamente a mesma frase pode querer dizer o oposto
  - *Está calor aquí*
  - *Gosto imenso da comida portuguesa*
- A LN presuppõe um interlocutor
- diálogo, jogo de perceber, explicar melhor, mudar a meio, deixar coisas implícitas
- jogo: regras, batota, combinar várias coisas, mudar de assunto, seguir estratégias, ter um objectivo

## Evolução e criatividade

- As línguas evoluem: o que é que isto **quer dizer?**
  - viubar -> enviar
  - conheço -> reconheço
  - costumar-se -> acostumar-se
  - amar hei-de -> amarei
  - casos -> artigos
  - amara -> tinha amado
  - buscar -> ir buscar
  - jogar a -> jogar à
  - amar a -> amar

## Referência a conhecimento implícito

- há um conjunto de conhecimentos que acomoda ("acama") uma interacção
- nunca se começa do nada (a não ser com bebés)
- *background*: cenário, ambiente, contexto
- *foreground*: à boca de cena, enfatizado, em destaque
- sem saber nada como é que se percebe um texto?
- com bastante conhecimento comum implícito é que a tradução (automática) funciona (Schubert)
- sem conhecimento comum não é possível a tradução (Quine)
- cultura é exactamente sobre o que se deixa implícito ou explícito

## Conhecimento implícito, cont.

- No matter how large our corpus, if it is domain specific, the major part of the domain ontology will not be specified because it is taken as given, or assumed to be part of the background knowledge the reader brings to the text. (Brewster et al., 2003)
- A text is an act of knowledge maintenance. (Brewster et al., 2005)
- a primary purpose of a text at some level is to change the relationship between existing concepts, or change the instantiations of these concepts [...] or adding new concepts to the existing domain ontology

## Vagueza: a propriedade mais importante

- A mesma unidade significa mais de uma coisa relacionada, e pode fazê-lo ao mesmo tempo.
  - *João é amigo do Pedro* amigo N ou ADJ? (categorização gramatical)
  - *Apaixonado, recusou o convite* (estar ou ser?), *Conhecer pessoas como ele é* uma aventura (conhecer bem, ou encontrar?)
  - *Passando por casa dela, lembrei-me do irmão* (temporal ou causal?)
  - *Encontraram-se na praia* (um ao outro, ou ambos lá?)
  - *A porta abriu-se!* (sozinha, ou alguém a abriu?)
  - *Ele quer casar com uma rapariga bonita* (que é, ou que seja?)
  - *O homem que matou X é louco* (de re ou de facto?)
  - *Ontem encontrei dois casais com o mesmo problema* (juntos ou separados?)
  - *Todos nesta sala sabem duas línguas* (as mesmas ou diferentes?)

## Vagueza, continuado: relação com o contexto

- Sem contexto não se pode compreender uma parte (palavra, expressão, frase)
- *A construção levou muito tempo*
- *A construção ficou muito bonita*
- *A construção deu-me imenso gozo; especializou-se em construção*
- *A polícia apanhou os ladrões no largo em Coimbra*
- *Conheço o JJ há cinco anos*
- *Conheci o JJ há cinco anos*
- *Toma arsénico, João, toma arsénico!*

## Outras propriedades “ilógicas”

- *Quando é que deixaste de bater na tua mulher?*
- *O rei de França é careca.*
- *Nunca comprei umas calças nos saldos sem que não me tenha arrependido*
- *Nunca comprei umas calças nos saldos sem que me tenha arrependido*
- *Não gosto de desporto algum.*
- *Não gosto de desporto nenhum*
- *Não gosto de nenhum desporto.*
- Metonímia: a parte pelo todo
- Opostos graduáveis: *um pequeno elefante > uma grande formiga*

## Irreferência

- *O presidente dos Estados Unidos vai encontrar-se com o primeiro ministro do Canadá em Março de 2007*
- *Quero comprar batatas fritas e comê-las com maionese*
- *O João perdeu o telemóvel ontem e a Maria pensa que o encontrou*
- *Os leões são perigosos (V)*
- *50% das pessoas sofrem de diabetes*
- *tenho comido muitos queques ultimamente (V)*
- *como ao meio dia (V)*
- *o fato de banho está algures no sótão*
- *quem quer que venha será recebido a tiro*
- *O João saiu antes de alguém chegar*

## Tipos de aplicações em PLN

- LN->LN: Extrair algo
- LN->X->LN: Arrumar algo para subsequente organização
- X->LN: Ajudar a criar algo
- Extra: descobrir plágio, descobrir um autor, autenticação, etc.
  
- tarefas com sentido
  - RAP, EI, TA
- tecnologias
  - análise sintáctica, POS tagging,...
- recursos e métodos
  - dicionários, ontologias, aprendizagem automática, prospecção (data mining), ...

## What is natural language (processing)?

- Natural language is the oldest and most successful knowledge representation language
- Used for communication, negotiation, and reason (->logic)
- Main features:
  - vagueness
  - context-dependent
  - implicit knowledge
  - evolves/dynamic/creative
- Different natural languages
  - different world view
  - different glue/implicit

## What is NL processing?

- Using computers to do things with natural language
- to be useful for humans
- Most intelligent human tasks involve language
  - as center (communicating, teaching, converting)
  - as periphery (mathematics papers, medical diagnosis)
- Daily tasks
  - writing (and creating or conveying information or affection)
  - reading (and finding information)
  - translating (and mediating)
  - teaching and learning and documenting
- Enormous political impact