LABORATOIRE D'INFORMATIQUE GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

27 de agosto de 2015
Universidade federal de São Carlos

Problemas de cientificidade na descrição da gramática e do léxico

Éric Laporte

Exemplo

The Science of Linguistics. *Inference. International Review of Science* 1(2), 2015.

L'avion s'est écrasé en mer
O avião caiu no mar

?*Le pilote a écrasé l'avion en mer*
≠O piloto (deixou + fez) o avião cair no mar
?O piloto amassou o avião no mar

Nenhuma descrição dá informação para todas as construções
Tem fatos demais

Adotar uma atitude científica
Confiabilidade, verificabilidade...
Maurice Gross, a partir dos anos 1960, introduziu práticas para tornar a descrição linguística mais científica
Hoje, a descrição linguística parece precisar disso mais ainda


20
36
3
29
11
31
2
4
17
18
1
1
Tato Cunha Photo

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A qualidade das respostas

O adjetivo *susceptible* "suscetível" aceita sujeito não humano?

Informação fornecida em tabelas (Picabia, 1976): não aceita

Rose est susceptible de devenir mère
Rose é suscetível de se tornar uma mãe

Le cas est susceptible de se produire
O caso é suscetível de acontecer

Informação errada

Informação fornecida

Conclusões coerentes com os dados

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A qualidade das perguntas

O adjetivo *susceptible* "suscetível" descreve a relação de um sujeito com a realização de uma ação?

Léger (2010): descreve

Paul est susceptible de tomber
Paul é suscetível de cair

La couleur est susceptible d'être légèrement différente
A cor é suscetível de ser levemente diferente

Pergunta imprecisa

É impossível avaliar se a respota é certa

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sumário

A arte de definir

Introspecção e corpus

O tamanho importa

Confrontação com a realidade

O linguista de poltrona

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A arte de definir

Critérios formais (Bloomfield, 1933)

Nomes contáveis

Tem muito leite

**Tem muita pessoa* (na norma culta)

A intuição semântica não é confiável

Análise distribucional

As precauções metodológicas das ciências experimentais

Estudo da linguagem ou do pensamento?

Protocolos experimentais (Gross, 1975)

**Le pilote a écrasé l'avion en mer*

Il a écrasé l'ail

Amassou o alho

s'écraser numa classe de verbos intransitivos

écraser numa classe de verbos transitivos

Léxico-gramática

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A arte de definir

Linguística de corpus

Análise distribucional: um adjuvante da intuição

Gramática de construções

Herdou da gramática gerativa a ideia de que procedimentos operacionais são inúteis (Chomsky, 1965)

Poucos linguistas hoje coletam dados empíricos de um modo formal e científico

Ainda é preciso coletar

Existe outra ciência em que alguém preconiza o abandono de precauções metodológicas, sem substituí-las por outras, e é respeitado?

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Subjetividade reprodutível

O conhecimento depende da intuição dos falantes

Tem muito leite

**Tem muita pessoa*

Concordância entre falantes sobre esse teste

O teste reorienta a investigação para um alvo mais preciso e mais realista

O adjetivo *susceptible* "suscetível" descreve a relação de um sujeito com a realização de uma ação

O adjetivo *digne* "digno" não descreve a relação de um sujeito com a realização de uma ação (Léger, 2010)

Difícil concordância entre falantes sobre esse teste

Reprodutibilidade

Não podemos ficar acostumados com **"concordância entre juízes" baixa**

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

O léxico-gramática

Cuidado, treino e talento do observador

Multiplicar os observadores e os alvos
Desistir de definições precisas

As precauções metodológicas do léxico-gramática
Escolher perguntas claras e precisas
Comparar julgamentos independentes (sessões coletivas)
Publicar os resultados
Treinar
Examinar centenas de entradas lexicais

Outras precauções para reprodutibilidade (Bisang, 2011)
20, 30 informantes
10 variantes "anodinas" dos exemplos
Informações sociológicas sobre os informantes

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Aceitabilidade

Formas sem sentido:

**A ideia dorme para baixo*
**Que a ideia durma para baixo nada*
**A ideia dorme para baixo nadam*
**Dorme para baixo nadam*

Ela não tem provavelmente saído

**Ela não tem provavelmente saído*
(baseado em Ernst, 2009)

Os falantes **aceitam** uma frase se pode ser usada em algum contexto para transmitir informações
Avaliam uma probabilidade
O modelo mais realista é binário

Gramaticalidade

Por definição (Chomsky, 1957): formas sem sentido podem ser gramaticais, se têm prosódia, são fáceis de lembrar...
Irrealista; inútil para PLN
Em prática: os linguistas gerativos rejeitam frases sem sentido. Dizem "gramatical" mas usam aceitabilidade

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Avaliação diferencial dos sentidos

A decoradora enfeita a vitrina com minissaias claras
A decoradora enfeita a vitrina com sua minissaia clara

Pedro diverte Paulo 1) de propósito
2) *Paulo acha¨Pedro divertido*

Noção introduzida por Gross (1975) mas implícita nos procedimentos formais de Bloomfield

Tem muita grama
Tem cem gramas

Mudança de sentido

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Avaliação diferencial dos sentidos

*Nas ciências físicas, é bem conhecido que as avaliações **absolutas** de uma variável (como a temperatura) sempre levam a resultados bastante grosseiros, em comparação com as avaliações **diferenciais** da mesma variável. A situação parece ser a mesma em linguística no que diz respeito aos sentidos.* Gross (1975:391-392)

Tem muita grama
Tem cem gramas

A tecnologia indica imediatamente que *grama* ocorre com *muita* no singular

Mas a análise distribucional por introspecção não é ultrapassada

A tecnologia extrai exemplos, mas não avalia **mudanças de sentido**

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Avaliação diferencial dos sentidos

Karl collects waste in the markets	1) Karl (coleta + cata) lixo nas feiras 2) ?Karl coleciona lixo nas feiras

As operações sintáticas produzem resultados diferentes conforme o sentido

Karl collects the waste in the markets

1) Karl (coleta + cata) o lixo nas feiras
~~2) Karl coleciona o lixo nas feiras~~

Karl does the collection of waste in the markets

1) Karl faz a (coleta + cata) de lixo nas feiras
~~2) Karl faz a coleção de lixo nas feiras~~

Karl makes a collection of waste in the markets

~~1) Karl faz uma (coleta + cata) de lixo nas feiras~~
2) ?Karl faz uma coleção de lixo nas feiras

Fenômenos gramaticais, não só de plausibilidade
Conhecimento útil para tradução

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Os léxicos-gramáticas

O método
Introspecção
Grande escala
Aceitabilidade binária
Avaliação diferencial dos sentidos
Sessões coletivas
Refinamento das perguntas

Os resultados
Construções com verbo suporte: *prestar atenção a*
Expressões idiomáticas: *abrir mão de*
A linguística descritiva ficou mais perto de ser uma disciplina empírica

O impacto
Poucos linguistas tiraram proveito dessa experiência
Muitos linguistas
- ou dispensam a introspecção
- ou a usam sem precauções notáveis

Publicações em francês

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sumário

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Introspecção e corpus

L'avion s'est écrasé en mer
?*Le pilote a écrasé l'avion en mer*

Introspecção
Distinguir entre formas raras e formas não em uso
Explorar os limites das variações
Necessita precauções metodológicas

Corpus
Formas que podiam não ser reparadas

Um corpus pode comprovar inaceitabilidade?

**There was house*	ausente de um grande corpus
honorificabilize	presente só na ativa

Necessidade de forjar formas

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Introspecção e corpus

Rejeição a priori da introspecção

"É perfeitamente contraditório descrever intuições como dados empíricos" (Sampson, 1979)

Falta um posicionamento sobre as contribuições metodológicas de Gross

Polarização entre gramática gerativa e linguística de corpus

"Já que os gerativistas não conseguem usar a introspecção direito, é absolutamente impossível usá-la"

Um raciocínio cubista

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Introspecção reprodutível ou não

Admite gradação semântica:		correlação	Combina com o advérbio *muito*:	
jovem	sim		*jovem*	sim
morto	não		*morto*	não
propriedade semântica, difícil de definir e observar			propriedade formal, mais fácil de definir e observar	

Em geral, nesse tipo de correlação, a propriedade semântica é percebida intuitivamente como a **causa** da outra

Tal intuição não é suficiente, é preciso verificar

Predicado semântico com um argumento que denota várias entidades:		correlação	Predicado com um argumento obrigatoriamente no plural:	
colecionar	sim		*colecionar*	sim
casar	**sim**		*casar*	**não**

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Introspecção reprodutível ou não

Admite gradação semântica:	
jovem	sim
morto	não

propriedade semântica,
difícil de definir e observar

correlação

Combina com o advérbio *muito*:	
jovem	sim
morto	não

propriedade formal,
mais fácil de definir e observar

Em geral, nesse tipo de correlação, a propriedade semântica é percebida intuitivamente como **mais importante** que a porpriedade formal

Assim mesmo, a propriedade formal pode ser

- **mais fácil de definir e observar**
- e **mais útil** para o PLN

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sumário

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Descrição sistemática do léxico

Cobertura lexical

A proporção do vocabulário de um idioma levada em conta num projeto

É difícil porque o léxico é **enorme** e **caótico**

Um trabalho gratificante

A gramática necessita descrição sistemática do léxico

Regras de gramática são generalizações a partir de um léxico

É preciso estudar o léxico para saber se uma regra é geral

O léxico-gramática do francês tem mais entradas do que qualquer grande projeto de dicionário do francês ou do inglês para o PLN: FrameNet, VerbNet, ComLex, e Meaning-Text.

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Vieses e objetividade

Em linguística, a **objetividade** significa que o linguista e os informantes sejam **distintos**

Evita **vieses psicológicos**

Neurolinguistas e psicolinguistas preferem experiências que garantem objetividade

Objetividade ≠ reprodutibilidade

Com protocolos linguísticos, resultados subjetivos podem ser reprodutíveis

Testes rotineiros de avaliação diferencial de sentidos necessitam que o informante seja um linguista treinado:

João fez um estudo do caso	=	*João estudou o caso*
João resumou um estudo do caso	≠	*João estudou o caso*

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Vieses e objetividade

Em linguística, **objetividade é realista?**

Atribuir categorias gramaticais
220 000 lemas
x 20 informantes
x vários contextos por palavra (ex. *regular*)
= dizenas de milhões de experiências objetivas

Um item lexical entra numa construção sintática?
Dezenas de milhares de palavras
20 informantes
Centenas de construções
A decisão necessita avaliação diferencial de sentidos
Resultado: é impossível descrever a gramática ou o léxico

Existem formas mais realistas de evitar vieses psicológicos
Controle por pares em sessões coletivas ou depois da publicação dos resultados

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sumário

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Corpus são necessários?

L'avion s'est écrasé en mer
?*Le pilote a écrasé l'avion en mer*

Os corpus garantem que os fatos são autênticos

Autenticidade

Se fatos autênticos são fatos empíricos confiáveis, voltamos à necessidade inicial

Se autenticidade exclui manipulação e introspecção, a exigência é contraproducente

Objetividade

É uma forma de garantir a reprodutibilidade

Não é a única

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Confrontação com a realidade

A aventura do léxico-gramática

Grande cobertura lexical e gramatical

Alto nível de detalhamento

Diferenças inesperadas entre itens lexicais

Comportamento sintático inesperadamente complexo

Discrepâncias inesperadas entre forma e sentido

A linguística de corpus

Os dois sentidos de *irritar* podem ser descritos em duas entradas lexicais se um aceita completiva sujeito e outro não

Um corpus ajuda a decidir só se os sentidos e as construções são representados

Produzir uma lista de entradas com propriedades necessita mais confrontação com a realidade

O vídeo irritou os radicais
O sabão irrita os olhos

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Expressões idiomáticas

O léxico-gramática das expressões idiomáticas
Começou como um sub-produto do inventário dos sentidos dos verbos (1968-1984)
Tabelas de propriedades sintáticas e semânticas

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Granularidade

Écraser l'ail et le sel	Amasse o alho e o sal
Tu m'écrases le pied	Você está pisando no meu pé
L'avion s'est écrasé en mer	O avião caiu no mar

+ 13 outras entradas de *écraser*

Divisão em entradas lexicais

Discretização do campo semântico

Em um sistema formal, uma propriedade deve ser propriedade de alguma coisa

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Granularidade

No léxico-gramática

Qualquer distinção estritamente correlada com uma propriedade reprodutivelmente observável é formalizada como separação de entradas

Exceção se for uma simples questão de transformação sintática

Écraser l'ail et le sel	*Écraser l'ail et le sel en purée*	Amasse o alho e o sal em um purê
Tu m'écrases le pied	**Tu m'écrases le pied en N*	*Você está pisando no meu pé em *N*

Um nível de granularidade satisfatório

Se a granularidade fosse mais fina: distinções não confiáveis

Mais grossa: atribuição de uma propriedade a um sentido que não a tem

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A intersecção entre sintaxe e léxico

CLASSE PB-TNH1																									
Freq. Absoluta Npred	Npred	Arg	N0=:Hum	N0=:nHum	N0=:Npc	N0=:Npred_de_N	N0=:QueF	Det=:E	Det=:ArtDef	Det=:um+Modif	GN	Simetria	N-a	N-v	Vsup=:ser_de	Vsup=:estar_com	Vsup=:fazer	Vsup=:sofrer	Vsup=:dar	Standard	Conversão	Ter_conversa	PI_obr	Nasp	Exemplo
3720	abertura	1	-	+	-	-	-	+	-	+	+	-	+	+	-	-	-	+	-	+	-	-	-	-	O recipiente tem uma abertura de 5mm.
6238	administração	1	-	+	-	-	-	-	-	+	+	-	-	+	-	+	-	-	-	+	-	-	-	-	O condomínio teve uma boa administração.
2950	aplicação	1	-	+	-	-	-	-	-	+	+	-	+	+	+	+	-	-	-	+	-	-	-	princípio	A teoria tem aplicação incontestável.
11443	aumento	1	-	+	-	+	-	-	-	+	+	-	-	+	-	+	-	-	-	+	-	-	-	-	O preço da uva teve um aumento de 25%.
2413	começo	1	-	+	-	-	-	-	-	+	+	-	-	+	-	-	-	-	-	+	-	-	-	-	O filme tem um começo ruim.
4678	construção	1	-	+	-	-	-	-	-	+	+	-	+	+	-	-	-	+	-	+	-	-	-	-	O prédio tem uma construção sólida.
6458	controle	1	-	+	-	-	-	+	-	+	+	-	+	+	-	-	-	-	-	+	-	-	-	-	O sistema tem um controle manual.
2613	cor	1	-	+	-	-	-	-	-	+	+	-	-	-	+	+	-	-	-	+	-	-	-	-	O mar tem uma cor incrível.
6113	corpo	1	-	+	-	-	-	-	-	+	+	-	+	-	+	+	-	-	-	+	-	-	-	-	O projeto tem um corpo bem definido.

Fonte: Santos, 2015

No cruzamento de duas listas

As entradas lexicais

As propriedades gramaticais

O formato tabelas

Natural, claro e legível para o trabalho descritivo

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A intersecção entre sintaxe e léxico

O formato cláusulas

Para análise sintática automática

Cada entrada é representada por uma ou várias fórmulas que especificam as propriedades positivas

As propriedades negativas ficam implícitas

O formato regras

Os linguistas computacionais são acostumados a regras

O formato tabelas é melhor

- Se as regras são verificadas antes de ser usadas, isso necessita uma codificação detalhada, do tipo de tabelas
- Se não são verificadas, não são confiáveis

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A intersecção entre sintaxe e léxico

book/books
ox/oxen
sheep/sheep
goose/geese

"Todas as gramáticas vazam" (Sapir, 1921)

Regras gerais de gramática sempre deixam escapar exceções

O léxico-gramática comprovou que a irregularidade em sintaxe é maciça

O estudo da sintaxe necessita o léxico

João está com gripe `ter(João, gripe)`?
`gripe(João)`?

Listando os nomes de doenças, Labelle (1986) reparou que a solução `gripe(João)` é melhor

João está com verruga no pé `ter(João, verruga, pé)`?
`verruga(João, pé)`?

O predicado `ter()` ia ter às vezes 2 argumentos, as vezes 3

O estudo do léxico necessita a sintaxe

Para separar entradas lexicais

Écraser l'ail et le sel	*Écraser l'ail et le sel en purée*	Amasse o alho e o sal em um purê
Tu m'écrases le pied	**Tu m'écrases le pied en N*	*Você está pisando no meu pé em *N*

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

A intersecção entre sintaxe e léxico

Segundo Chomsky (1970), quanto mais se estuda léxico, tanto menos se estuda sintaxe, e vice-versa

Muitos linguistas gerativos acreditam que **não existe** intersecção entre sintaxe e léxico

Muitos linguistas acreditam que a sintaxe é mais científica e mais teórica do que o léxico

"Estudar o léxico é catar conchas na praia" (Silver, 1998)

Existe outra ciência em que seguem alguém que desprestigia a coleta de fatos empíricos?

Observando estrelas uma por uma, Hubble (1925) descobriu as galáxias

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sumário

A arte de definir

Introspecção e corpus

O tamanho importa

Confrontação com a realidade

O linguista de poltrona

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

O linguista de poltrona

A impostura de Chomsky

Só importa o que reflete processos gramaticais regulares

Explorar dados empíricos não serve

Procedimentos formais não servem

A contrarrevolução da linguística de corpus

Pelo direito a dados empíricos

O "linguista de poltrona" (Fillmore) caricatura a falta de cuidado na observação e de confrontação com a realidade linguística

Reabilitou o uso de corpus

Não reabilitou:

- os procedimentos formais de observação empírica
- o estudo sistemárico do léxico

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Conclusão

Uma linguística descritiva científica é possível

Rigorosa

Realista

Útil para sistemas híbridos de PLN

Uma história movimentada

O léxico-gramática acabou quase desconhecido dos linguistas e do PLN

Existe um legado de recursos e de publicações

Um ponto de partida prometedor