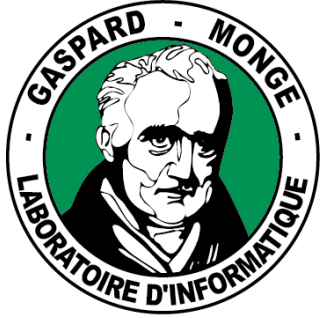




**LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE**

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

16/11/2017

CONEL

**Universidade federal do
Espírito Santo**

Como reunir dados linguísticos cientificamente

Eric Laporte



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

Precisão

Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

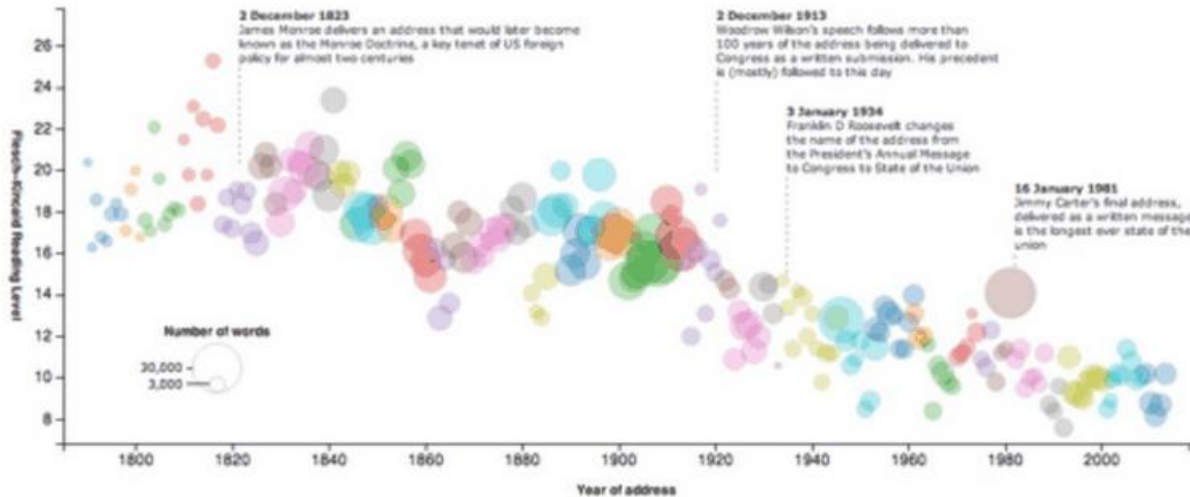
Reprodutibilidade



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Nossos resultados são válidos?



Fonte: *The Guardian*

Nossa pesquisa em linguística usa dados

O valor dos resultados depende do valor dos dados

Os resultados concordam com a realidade?

Consequências na sociedade

- uso na justiça

- uso no computador

...



Reunir dados cientificamente



By brian0918™ (Own work) [Public domain], via
Wikimedia Commons

- Um caminho para produzir resultados confiáveis é reunir dados cientificamente
- Existem outros caminhos
- É nossa decisão escolhermos o caminho que preferirmos
- Só sugiro não decidirmos por más razões ligadas ao prestígio da ciência, vista como dominante:
- para nos aproveitar deste prestígio
 - por rejeitar a ciência por ocupar injustamente uma posição dominante
 - por ficar constrangidos, com medo de não dar conta

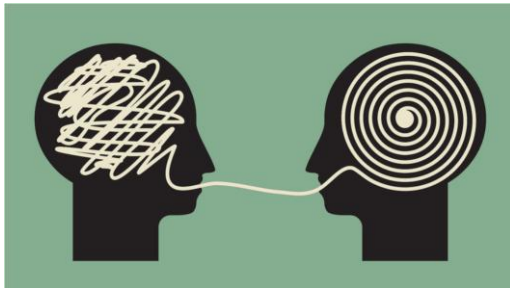


LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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



The original uploader was Louisehui at English Wikipedia
(Transferred from en.wikipedia to Commons.)



The Big Think



Quais as características da ciência?

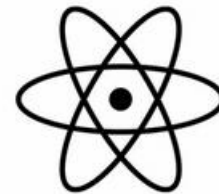
Alguns estereótipos...

Ciência e tecnologia

Ciência, seriedade, racionalidade

Ciência e objetividade

Vamos utilizar referências em outras disciplinas





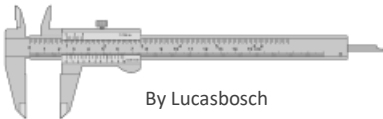
LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Precisão



This image was created by Booyabazooka (Based on Image:Rubiks cube.jpg)



By Lucasbosch

Em matemática

Precisão absoluta

Em física

Medidas incluem uma avaliação da possibilidade de erro

E em linguística?



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Precisão em semântica



By Evan-Amos (Own work) [Public domain], via
Wikimedia Commons



By Olivenöl_Weißbrot_und_grobes_Salz.JPG:
4028mdk09 derivative work: Andrzej 22

Muitas categorias são imprecisas
Contável/incontável
ideia, pensamento



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Precisão em linguística descritiva

Entendeu

Duvidou da notícia



Verbo transitivo/intransitivo: uma categoria imprecisa

Uso intransitivo de um verbo transitivo

Verbo transitivo indireto

Crowdsourcing, descoberta coletiva de conhecimento

Custo baixo

Não tem uma razão pela qual os resultados são conformes à realidade

Uma má resposta ao problema das categorias imprecisas

Construções

transitiva direta / transitiva indireta / intransitiva / ...

Subestrutura de uma construção

Adotar definições mais precisas

Ciência não quer dizer tecnologia

É mais uma questão de método de descoberta

Entendeu o princípio

Duvidou da notícia

Arrasou



Entendeu

Duvidou



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

Precisão

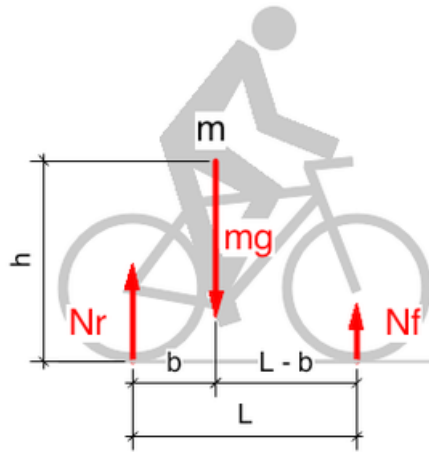
Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

Reprodutibilidade



Modelos

Um modelo é uma visão simplificada de um sistema ou de um processo

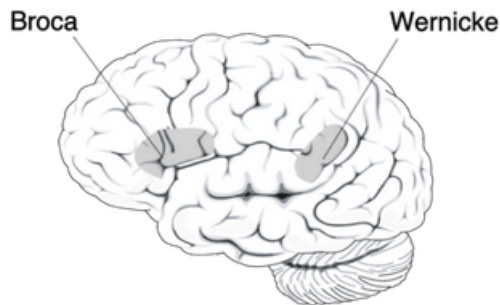
Em dinâmica

Ciência das forças e dos movimentos

Em muitos contextos, pode-se negligenciar a influência da rotação da Terra



Modelos em psicolinguística



Selecionar alvos de pesquisa

Omitir alvos inacessíveis

Os detalhes da representação de regras sintáticas na memória do falante são inacessíveis

Duas construções similares são armazenadas separadamente ou com uma parte em comum?

Entendeu o princípio

Duvidou da notícia

Arrasou

Aproximou o cartão do leitor

Entendeu

Duvidou

Aproximou o cartão



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Modelos em linguística descritiva



By Evan-Amos (Own work) [Public domain], via
Wikimedia Commons

Selecionar alvos de pesquisa

Comportamento **sintático-semântico** de nome
contável/incontável

Quantos ovos usa?

Quantos tipos de ovos usa?

} diferença de sentido

Quantos azeites usa?

Quantos tipos de azeites usa?

} conservação do sentido



Produz resultados relevantes para a categoria **semântica**
contável/incontável

By Olivenöl_Weißbrot_und_grobes_Salz.JPG:
4028mdk09 derivative work: Andrzej 22



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

Precisão

Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

Reprodutibilidade



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Verificabilidade

Em jornalismo

Não é uma ciência, mas a verificação é essencial

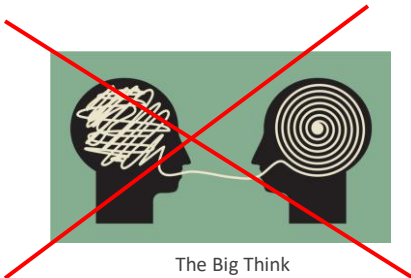
Em psicanálise e psiquiatria

A intuição sobre o fato de que o sujeito sofra ou não é uma base da prática

Não se pode **verificar** se o sujeito sofre

Esse ramo não deixa de ser sério, racional e útil à sociedade

Ciência não quer dizer seriedade nem racionalidade

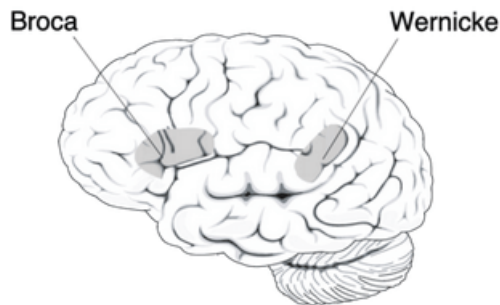




LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Verificabilidade em psicolinguística



Podemos fazer hipóteses sobre os detalhes da representação
de regras sintáticas na memória do falante

Tais hipóteses são inverificáveis



Verificabilidade em gramática gerativa

Uma gramática é uma representação matemática da sintaxe
e da semântica de uma língua

Determinar se uma gramática é melhor do que outra

Supondo que ambas gramáticas são conformes aos fatos

Preferir a com o menor número de símbolos

Inaplicável enquanto ambas descrições não estiverem
completas



Falsificabilidade

Um tipo de resultado é falsificável se conhecemos uma forma de comprovar que o resultado está errado (Karl Popper, 1959)

"A noção de sílaba existe em qualquer língua"

Falsificável (achando uma língua em que não exista)

A falsificabilidade é menos exigente do que a verificabilidade

Não verificável (só estudando todas as línguas uma por uma)



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

Precisão

Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

Reprodutibilidade



Como reunir dados llingüísticos científicamente • 19

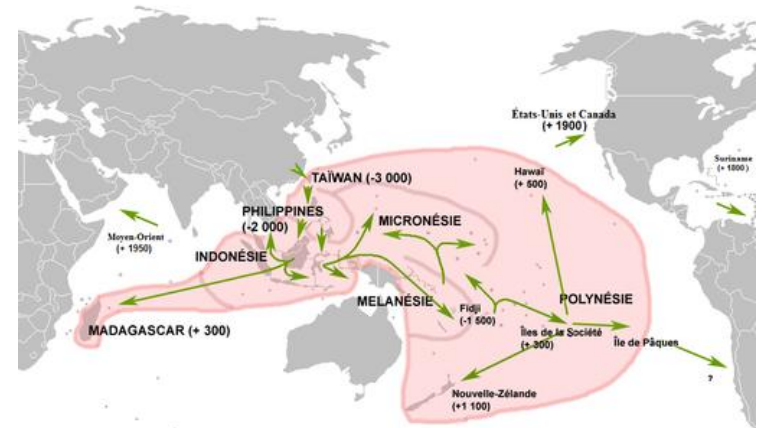


Dados em linguística histórica

Italian	Spanish	Portuguese	French
corpo	cuerpo	corpo	corps
crudo	crudo	cru	cru
catena	cadena	cadeia	chaîne
cacciare	cazar	caçar	chasser

Fonte: Wikipédia em inglês

Fontes escritas
Leis comparatistas



By English: Maulucioni, based on previous work by Christophe Cagé



Dados em sociolinguística

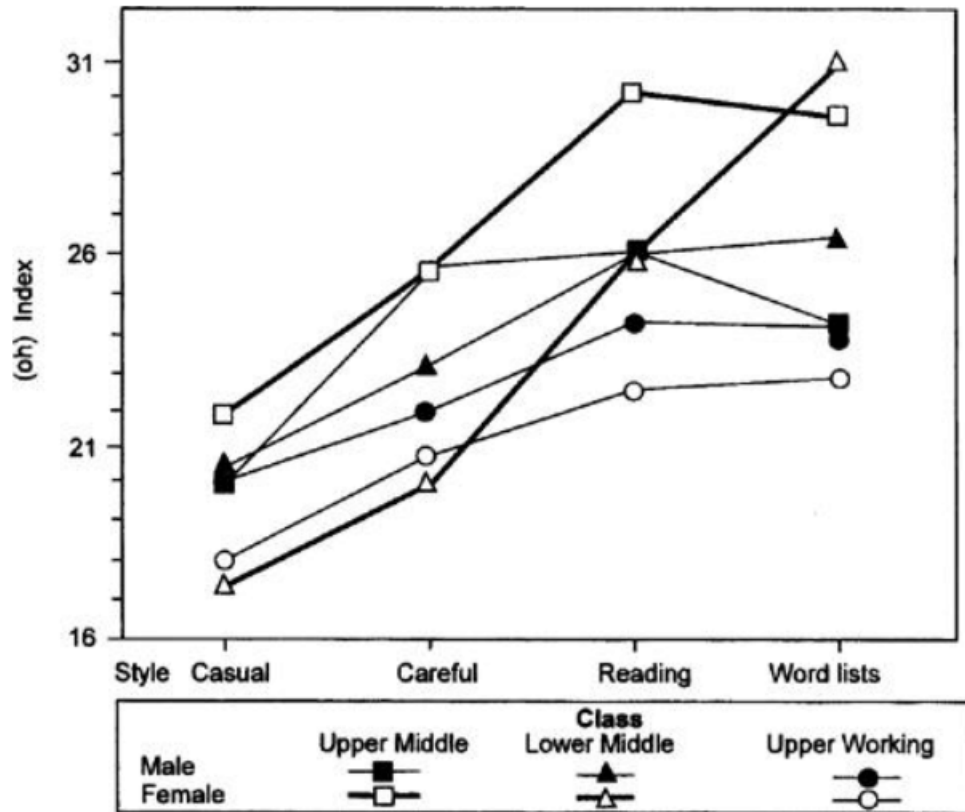


Figure 8.1 Style shifting of (oh) by three socioeconomic groups in New York City (after Labov, 1990: 224)

Amostra de casos espalhados em um espaço de variação

Parâmetros

Sexo, classe social, situação de comunicação

Idade, origem geográfica, nível de educação



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

Precisão

Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

Reprodutibilidade



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Protocolos experimentais

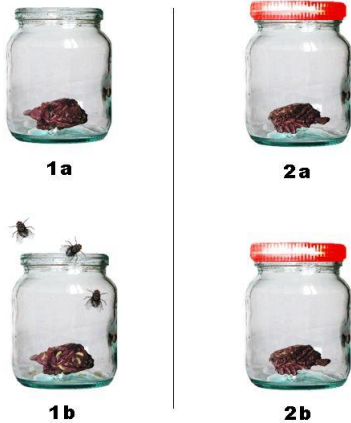


By Jojojpjohoilo (Own work) [CC BY-SA 4.0
(<https://creativecommons.org/licenses/by-sa/4.0/>)],
via Wikimedia Commons

Em biologia

Geração espontânea significa o surgimento de seres vivos a partir de matéria inanimada

Francesco Redi (1668) realizou uma experiência que comprovou que carne podre protegida por gaza não gera larvas de moscas



By Aushulz [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY-SA 3.0
(<https://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Protocolos experimentais em linguística de corpus

By Milos Jakubicek, acting on behalf of Lexical Computing (Own work) [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)], via Wikimedia Commons

Observar dados autênticos, isto é, produzidos
independentemente do pesquisador

Com este protocolo experimental, é impossível
observar que uma forma não está em uso

Fazer de conta é usado com complemento nominal?
A terra tremeu nominaliza?

Este protocolo não contribui à exploração a fronteira
da linguagem

Fez de conta de analisar a nota

**Fez de conta uma análise da nota*



Protocolos experimentais



Em biologia

A germinação de uma semente necessita

- umidade?
- oxigênio?
- uma certa faixa de temperatura?
- um certo nível de luz?

Experiência	Umidade	Oxigênio
1	+	+
2	+	-
3	-	+
4	-	-



Protocolos experimentais em sociolinguística

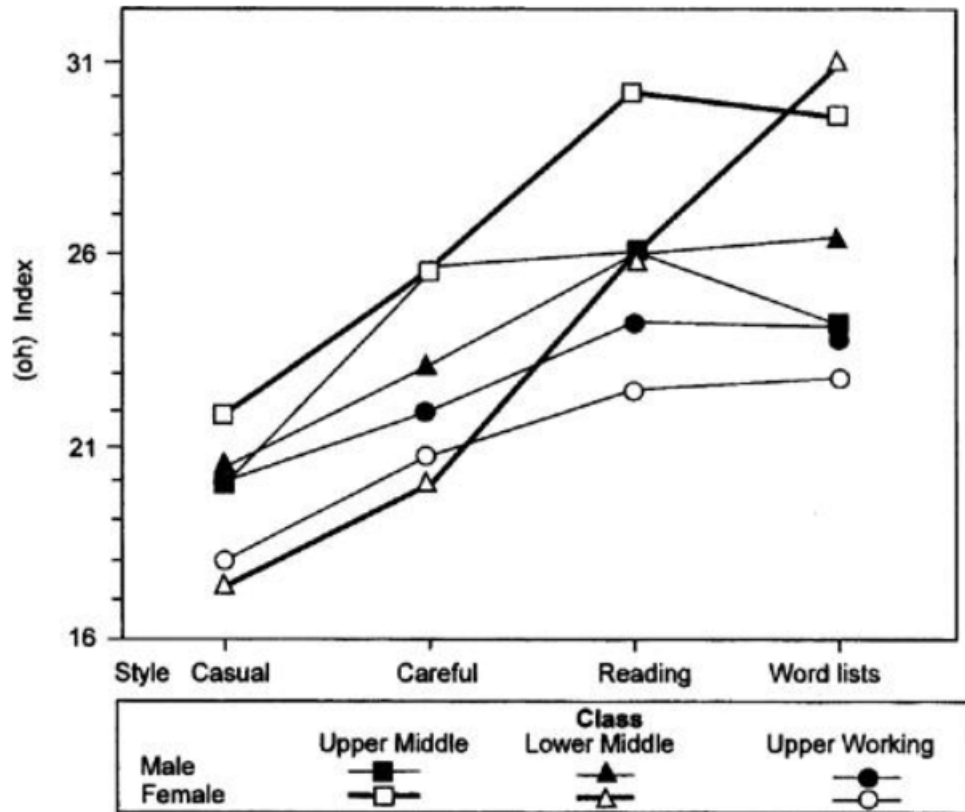


Figure 8.1 Style shifting of (oh) by three socioeconomic groups in New York City (after Labov, 1990: 224)

Observação de casos

Entrevista com elicitación

Gravação da própria fala

Observação em laboratório

Dificuldades

Naturalidade da situação de comunicação

Conhecimento do alvo da pesquisa



Protocolos experimentais em linguística descritiva

Protocolo experimental para observar o efeito de cada
parâmetro

**5 reais eram validos pelo sorvete*

Os 5 reais foram engordurados pelo sorvete

Diferenças entre as duas observações

Sujeito indefinido/definido

Tempo do verbo

Valor lexical do verbo

Não dá para concluir se a primeira sequência é incorreta

- porque o sujeito é indefinido

- porque o auxiliar está no imperfeito

- ou porque o verbo *valer* não tem passiva

**Os 5 reais eram validos pelo sorvete*

Os 5 reais eram engordurados pelo sorvete



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

Precisão

Modelos

Verificabilidade e falsificabilidade

Dados

Protocolos experimentais

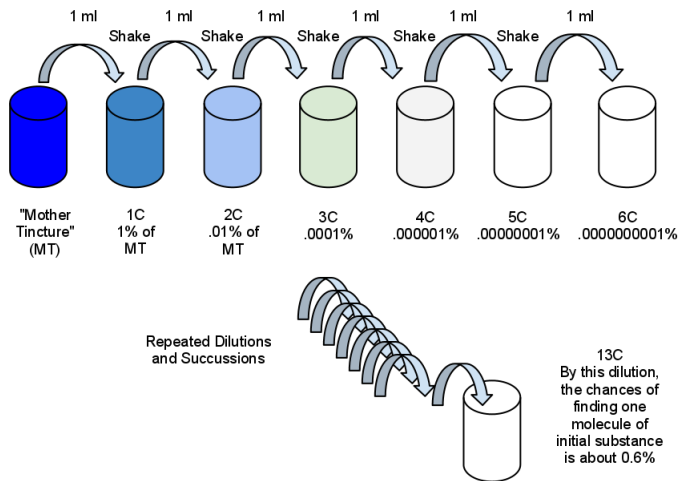
Reprodutibilidade



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Reprodutibilidade



Hahnemann's Method of Dilution and Succussion

Uma experiência é reprodutível se leva ao mesmo resultado sempre que é repetida

Em medicina

Jacques Benveniste (1988) evidenciou propriedades biológicas de proteínas diluídas na água, mesmo com uma diluição tamanha que não tem nenhuma molécula da proteína

A repetição da experiência não dá sempre esse mesmo resultado



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Reprodutibilidade e objetividade



Reprodutibilidade tem a ver com objetividade

Quando o experimentador tenta comprovar o que já pensa saber por intuição, pode errar

Reprodutibilidade e subjetividade

Uma experiência que usa subjetividade pode ser reprodutível

Repetir a experiência com outros sujeitos

*[É melhor não] privilegiarmos a aparência
do conhecimento em relação ao trabalho
de descobrirmos*

Ta-Nehisi Coates, 2017



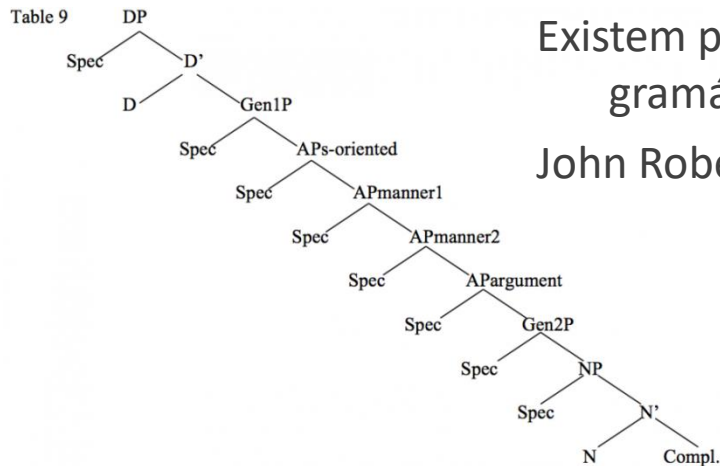
Reprodutibilidade em gramática gerativa

Segundo Noam Chomsky (1957), a intuição de
gramaticalidade é direta, como a intuição matemática

Não necessita protocolo experimental

Existem poucos trabalhos sobre reprodutibilidade em
gramática gerativa

John Robert Ross, 1979





Reprodutibilidade em linguística descritiva

Crîtérios formais

A observação de comportamentos sintático-semânticos numa comunidade linguística pode ser reprodutível

Estruturalismo americano

<i>Quantos ovos usa?</i>	}	diferença de sentido
<i>Quantos tipos de ovos usa?</i>		
<i>Quantos azeites usa?</i>	}	conservação do sentido
<i>Quantos tipos de azeites usa?</i>		

Um empreendimento de descrição linguística em grande escala desenvolveu um protocolo experimental (Maurice Gross, 1975, 1982)

- Forjar sequências para explorar os limites da linguagem
- Avaliar a aceitabilidade e as diferenças semânticas por introspecção
- Repetir a experiência: sessões coletivas, controle pelos pares



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

Nem sempre é difícil encontrar um método científico para
estudar um problema

Os métodos dependem muito do ramo de estudo



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Obrigado

CONTATO

ERIC LAPORTE

00 +33 (0)1 60 95 75 52

ERIC.LAPORTE@UNIV-PARIS-EST.FR