

# AULA EXTRA: DETECÇÃO AUTOMÁTICA DE NOTÍCIAS FALSAS

SCC0633 e SCC5908 Processamento de Linguagem Natural

Roney L. de S. Santos

[roneysantos@usp.br](mailto:roneysantos@usp.br)

Thiago A. S. Pardo

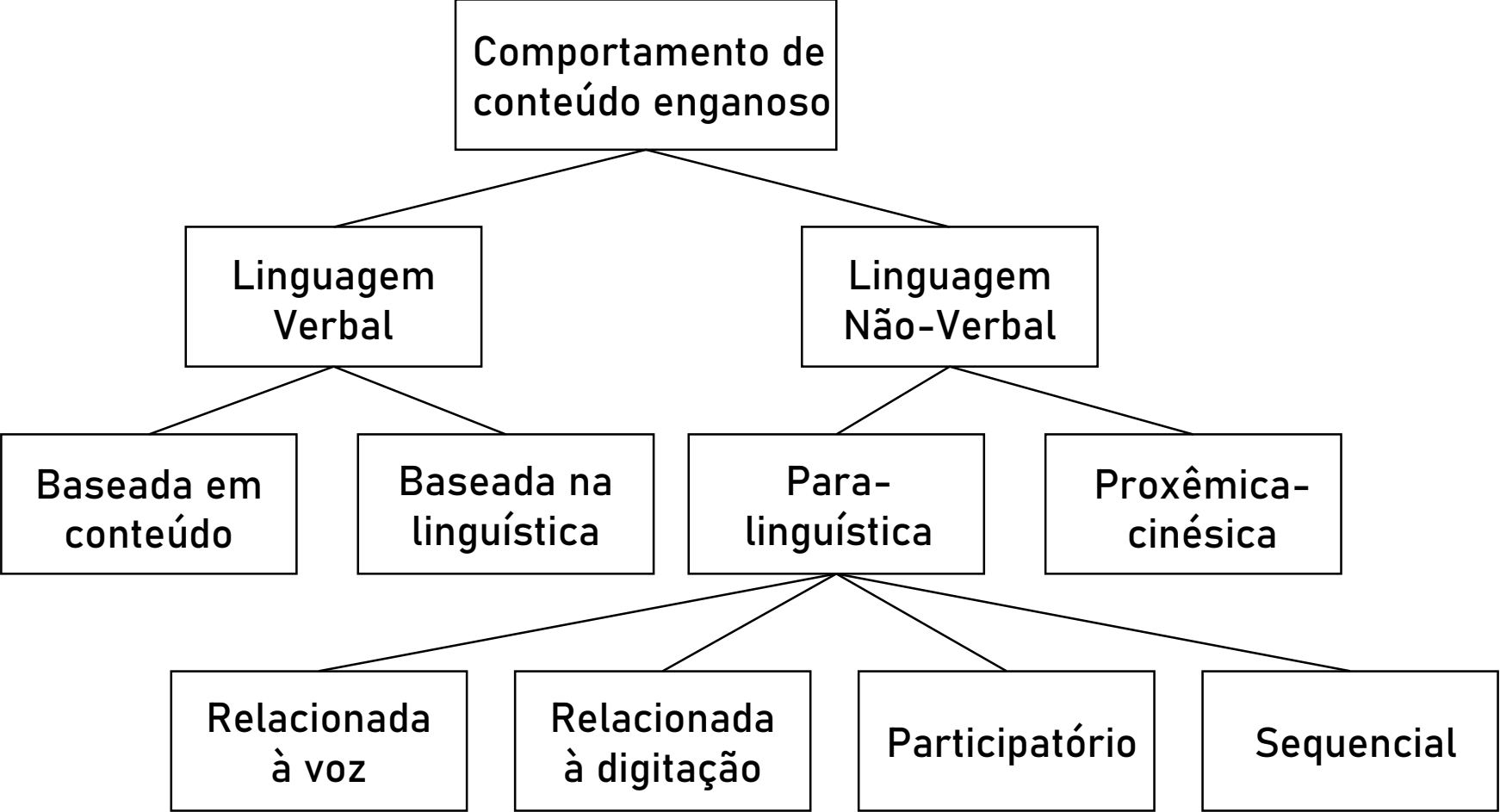
[taspardo@icmc.usp.br](mailto:taspardo@icmc.usp.br)

# CONTEXTO ATUAL

- Notícia falsa tem mais alcance que as verdadeiras nas redes sociais
  - Morte de famosos, notícias sobre política, o que não comer, o que não tomar...
  - E aí pode ter consequências terríveis, como na Índia...
- **Nem todo mundo segue as dicas** para não se enganar por fake news, muito presente na imprensa real.
- Que tal fazer uma **detecção e filtragem automática** dessas notícias?
- Abordagens baseadas na **língua** e no **conteúdo** do texto.
  - Hipótese de que o enganador não se preocupa com a linguagem expressada no texto falso.

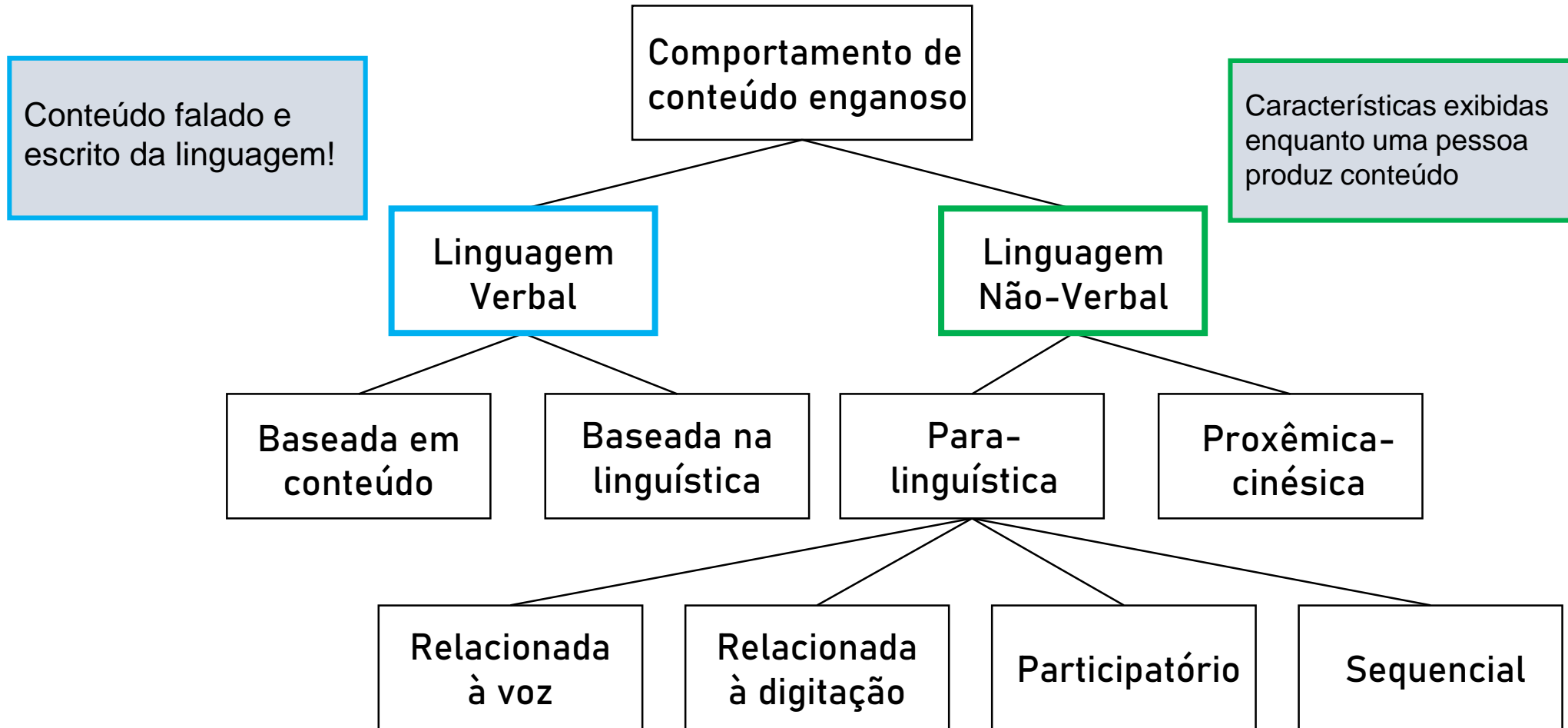
# TAXONOMIA DO COMPORTAMENTO DE CONTEÚDO ENGANOSO

Zhou, 2005



# TAXONOMIA DO COMPORTAMENTO DE CONTEÚDO ENGANOSO

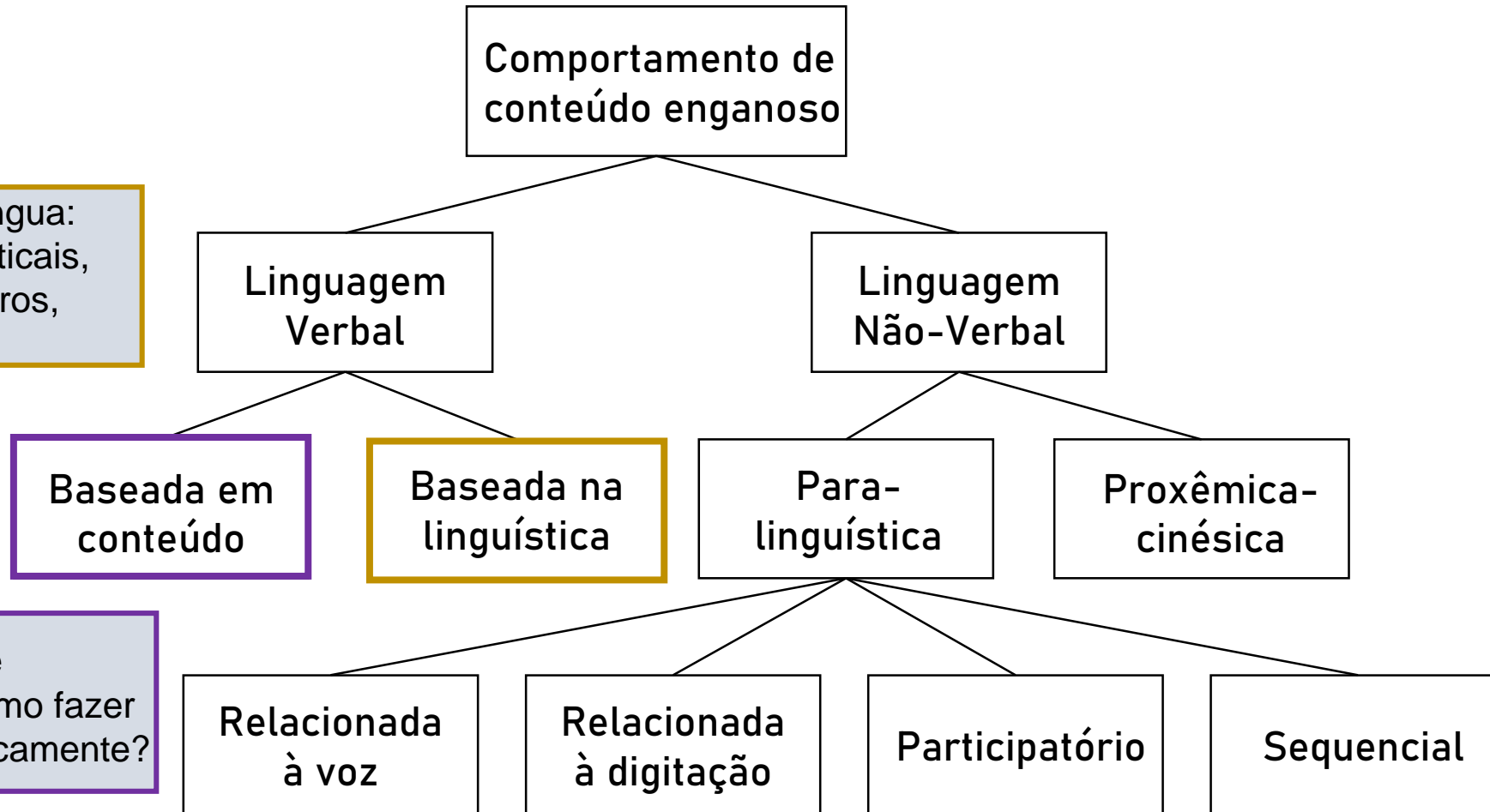
Zhou, 2005



# TAXONOMIA DO COMPORTAMENTO DE CONTEÚDO ENGANOSO

Zhou, 2005

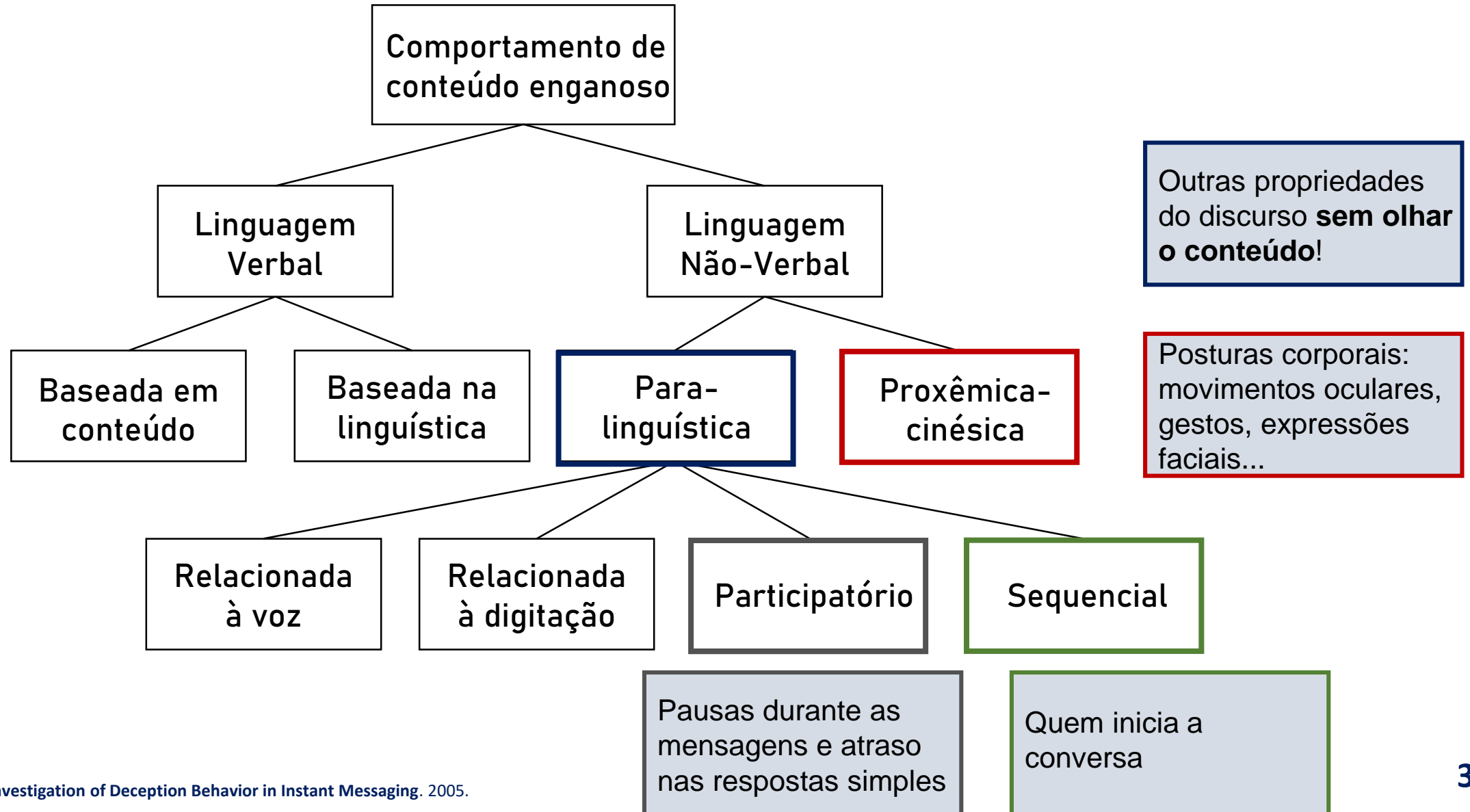
Atributos da língua:  
classes gramaticais,  
semânticas, erros,  
diversidade...



Checagem de conteúdo! Como fazer isso automaticamente?

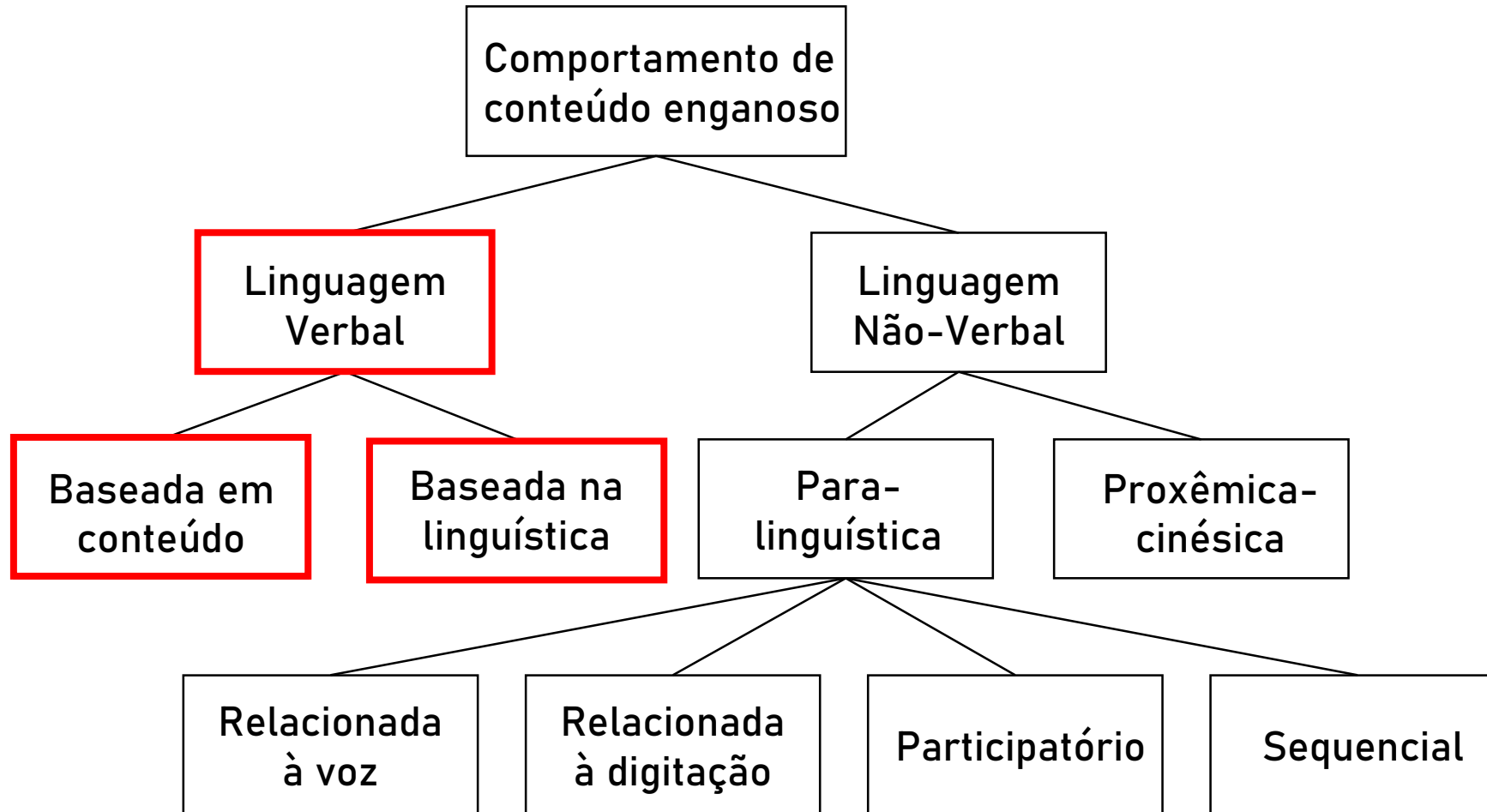
# TAXONOMIA DO COMPORTAMENTO DE CONTEÚDO ENGANOSO

Zhou, 2005



# TAXONOMIA DO COMPORTAMENTO DE CONTEÚDO ENGANOSO

Zhou, 2005



# EXEMPLOS DE ATRIBUTOS

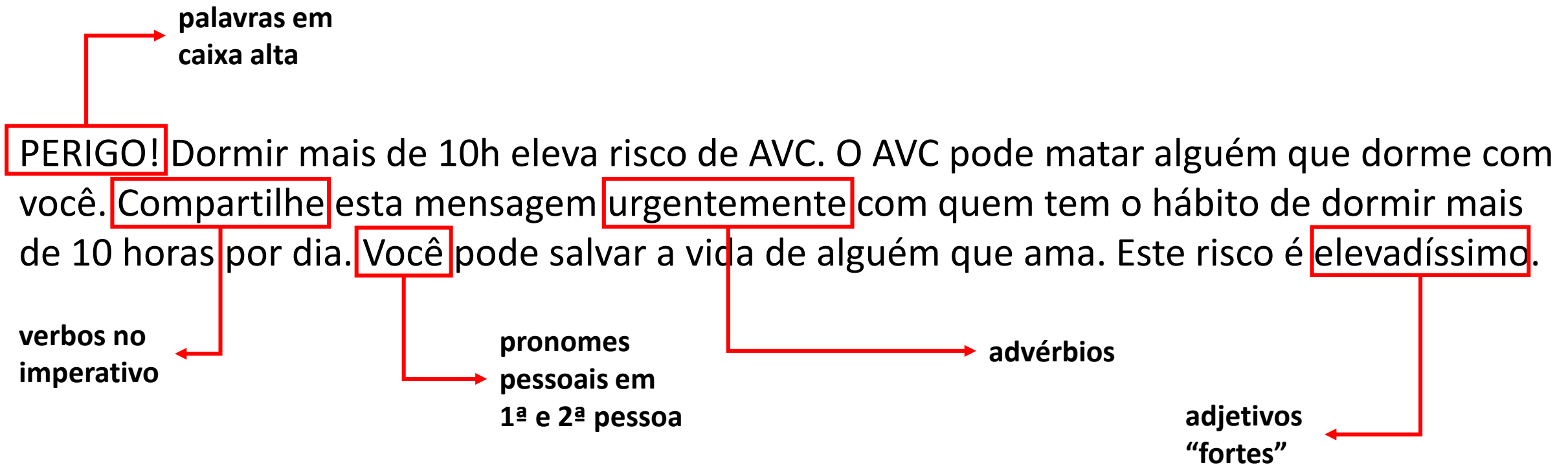
- CLASSES GRAMATICAIS
- CLASSES SEMÂNTICAS
- NÚMERO DE SENTENÇAS, PALAVRAS, PONTUAÇÃO...
- PALAVRAS ESCRITAS EM CAIXA ALTA
- ERROS ORTOGRÁFICOS
- DIVERSIDADE DE VOCABULÁRIO
- PAUSALIDADE
- EMOTIVIDADE
- NÃO-IMEDIATISMO
- INCERTEZA
- ...



# CARACTERÍSTICAS ENCONTRADAS NAS FAKE NEWS

PERIGO! Dormir mais de 10h eleva risco de AVC. O AVC pode matar alguém que dorme com você. Compartilhe esta mensagem urgentemente com quem tem o hábito de dormir mais de 10 horas por dia. Você pode salvar a vida de alguém que ama. Este risco é elevadíssimo.

# CARACTERÍSTICAS ENCONTRADAS NAS FAKE NEWS



# CARACTERÍSTICAS ENCONTRADAS NAS FAKE NEWS

Como encontramos essas características?

# CARACTERÍSTICAS ENCONTRADAS NAS FAKE NEWS

Como encontramos essas características?

Com o uso de **etiquetadores morfossintáticos** e semânticos!

**Normalizadores**

spaCy

97% de acurácia na detecção das classes gramaticais (lg)

PALAVRAS (Bick, 2000)

97% de acurácia na detecção das classes gramaticais

NLPNet (Fonseca et al., 2013)

97,33% de acurácia na detecção das classes gramaticais

ENELVO (Bertaglia e Nunes, 2016)

92% de acurácia na correção de erros ortográficos

# EXEMPLO DE SAÍDA DO PALAVRAS (BICK, 2000)

**PERIGO! Dormir mais de 10h eleva risco de AVC. O AVC pode matar alguém que dorme com você. Compartilhe esta mensagem urgentemente com quem tem o hábito de dormir mais de 10 horas por dia. Você pode salvar a vida de alguém que ama. Este risco é elevadíssimo.**

([('Perigo', {'lemma': 'perigo', 'morpho': ['N', 'M', 'S'], 'POS': 'NPHR', 'semantic': ['\*', 'ac', 'sit'], 'relation': '#1->0'}), ('!', {'lemma': '!', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#2->0'}), ('Dormir', {'lemma': 'dormir', 'morpho': ['V', 'INF'], 'POS': 'ICL-SUBJ>', 'semantic': ['\*', 'vH', 'mv'], 'relation': '#1001->1004'}), ('mais=de', {'lemma': 'mais=de', 'morpho': ['ADV', '@<ADVL'], 'POS': '<ADVL', 'semantic': ['quant', 'ADVL'], 'relation': '#1002->1001'}), ('10h', {'lemma': '10h', 'morpho': ['NUM', 'M/F', 'P'], 'POS': '<ADVL', 'semantic': ['card', 'ADVL'], 'relation': '#1003->1001'}), ('eleva', {'lemma': 'elevar', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-STA', 'semantic': ['vH', 'fmc', 'mv'], 'relation': '#1004->0'}), ('risco', {'lemma': 'risco', 'morpho': ['N', 'M', 'S'], 'POS': '<SUBJ', 'semantic': ['ac', 'ac-sign', 'SUBJ'], 'relation': '#1005->1004'}), ('de', {'lemma': 'de', 'morpho': ['PRP'], 'POS': 'N<', 'semantic': ['np-close'], 'relation': '#1006->1005'}), ('AVC', {'lemma': 'AVC', 'morpho': ['PROP', 'M/F', 'S'], 'POS': 'P<', 'semantic': ['org', '\*'], 'relation': '#1007->1006'}), ('.', {'lemma': '.', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#1008->0'}), ('O', {'lemma': 'o', 'morpho': ['DET', 'M', 'S'], 'POS': '>N', 'semantic': ['\*', 'artd'], 'relation': '#2001->2002'}), ('AVC', {'lemma': 'AVC', 'morpho': ['PROP', 'M', 'S'], 'POS': 'SUBJ>', 'semantic': ['inst', '\*'], 'relation': '#2002->2003'}), ('pode', {'lemma': 'poder', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-STA', 'semantic': ['fmc', 'aux'], 'relation': '#2003->0'}), ('matar', {'lemma': 'matar', 'morpho': ['V', 'INF'], 'POS': 'ICL-AUX<', 'semantic': ['mv'], 'relation': '#2004->2003'}), ('alguém', {'lemma': 'alguém', 'morpho': ['SPEC', 'M', 'S'], 'POS': '<ACC', 'semantic': ['ACC'], 'relation': '#2005->2004'}), ('que', {'lemma': 'que', 'morpho': ['SPEC', 'M', 'S'], 'POS': 'SUBJ>', 'semantic': ['clb', 'clb-fs', 'rel'], 'relation': '#2006->2007'}), ('dorme', {'lemma': 'dormir', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-N<', 'semantic': ['vH', 'mv', 'np-close'], 'relation': '#2007->2005'}), ('com', {'lemma': 'com', 'morpho': ['PRP'], 'POS': '<PIV', 'semantic': ['PIV'], 'relation': '#2008->2007'}), ('você', {'lemma': 'você', 'morpho': ['PERS', 'M/F', '3S', 'NOM/PIV'], 'POS': 'P<', 'semantic': [], 'relation': '#2009->2008'}), ('.', {'lemma': '.', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#2010->0'}), ('Compartilhe', {'lemma': 'compartilhar', 'morpho': ['V', 'PR', '1/3S', 'IMP', 'VFIN'], 'POS': 'FS-STA', 'semantic': ['\*', 'vH', 'fmc', 'mv'], 'relation': '#3001->0'}), ('esta', {'lemma': 'este', 'morpho': ['DET', 'F', 'S'], 'POS': '>N', 'semantic': ['dem'], 'relation': '#3002->3003'}), ('mensagem', {'lemma': 'mensagem', 'morpho': ['N', 'F', 'S'], 'POS': '<ACC', 'semantic': ['sem-s', 'ACC'], 'relation': '#3003->3001'}), ('urgentemente', {'lemma': 'urgentemente', 'morpho': ['ADV'], 'POS': '<ADVL', 'semantic': ['nh', 'ADVL'], 'relation': '#3004->3001'}), ('com', {'lemma': 'com', 'morpho': ['PRP'], 'POS': '<ADVL', 'semantic': ['ADVL'], 'relation': '#3005->3001'}), ('quem', {'lemma': 'quem', 'morpho': ['SPEC', 'M/F', 'S'], 'POS': 'SUBJ>', 'semantic': ['clb', 'clb-fs', 'rel'], 'relation': '#3006->3007'}), ('tem', {'lemma': 'ter', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-P<', 'semantic': ['mv'], 'relation': '#3007->3005'}), ('o', {'lemma': 'o', 'morpho': ['DET', 'M', 'S'], 'POS': '>N', 'semantic': ['artd'], 'relation': '#3008->3009'}), ('hábito', {'lemma': 'hábito', 'morpho': ['N', 'M', 'S'], 'POS': '<ACC', 'semantic': ['f-psych', 'percep-f', 'cloH', 'ACC'], 'relation': '#3009->3007'}), ('de', {'lemma': 'de', 'morpho': ['PRP'], 'POS': 'N<', 'semantic': ['np-close'], 'relation': '#3010->3009'}), ('dormir', {'lemma': 'dormir', 'morpho': ['V', 'INF'], 'POS': 'ICL-P<', 'semantic': ['vH', 'mv'], 'relation': '#3011->3010'}), ('mais=de', {'lemma': 'mais=de', 'morpho': ['ADV'], 'POS': '>A', 'semantic': ['quant'], 'relation': '#3012->3013'}), ('10', {'lemma': '10', 'morpho': ['NUM', 'F', 'P'], 'POS': '>N', 'semantic': ['card'], 'relation': '#3013->3014'}), ('horas', {'lemma': 'hora', 'morpho': ['N', 'F', 'P'], 'POS': '<ACC', 'semantic': ['dur', 'temp', 'unit', 'ACC'], 'relation': '#3014->3011'}), ('por', {'lemma': 'por', 'morpho': ['PRP'], 'POS': 'N<', 'semantic': ['np-close'], 'relation': '#3015->3014'}), ('dia', {'lemma': 'dia', 'morpho': ['N', 'M', 'S'], 'POS': 'P<', 'semantic': ['temp', 'dur', 'per', 'unit'], 'relation': '#3016->3015'}), ('.', {'lemma': '.', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#3017->0'}), ('Você', {'lemma': 'você', 'morpho': ['PERS', 'M/F', '2S', 'NOM'], 'POS': 'SUBJ>', 'semantic': ['\*'], 'relation': '#4001->4002'}), ('pode', {'lemma': 'poder', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-STA', 'semantic': ['fmc', 'aux'], 'relation': '#4002->0'}), ('salvar', {'lemma': 'salvar', 'morpho': ['V', 'INF'], 'POS': 'ICL-AUX<', 'semantic': ['vH', 'mv'], 'relation': '#4003->4002'}), ('a', {'lemma': 'o', 'morpho': ['DET', 'F', 'S'], 'POS': '>N', 'semantic': ['artd'], 'relation': '#4004->4005'}), ('vida', {'lemma': 'vida', 'morpho': ['N', 'F', 'S'], 'POS': '<ACC', 'semantic': ['per', 'ACC'], 'relation': '#4005->4003'}), ('de', {'lemma': 'de', 'morpho': ['PRP'], 'POS': '<PIV', 'semantic': ['PIV'], 'relation': '#4006->4003'}), ('alguém', {'lemma': 'alguém', 'morpho': ['SPEC', 'M', 'S'], 'POS': 'P<', 'semantic': [], 'relation': '#4007->4006'}), ('que', {'lemma': 'que', 'morpho': ['SPEC', 'M', 'S'], 'POS': 'ACC>', 'semantic': ['clb', 'clb-fs', 'rel'], 'relation': '#4008->4009'}), ('ama', {'lemma': 'amar', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-N<', 'semantic': ['vH', 'mv', 'np-close'], 'relation': '#4009->4007'}), ('.', {'lemma': '.', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#4010->0'}), ('Este', {'lemma': 'este', 'morpho': ['DET', 'M', 'S'], 'POS': '>N', 'semantic': ['\*', 'dem'], 'relation': '#5001->5002'}), ('risco', {'lemma': 'risco', 'morpho': ['N', 'M', 'S'], 'POS': 'SUBJ>', 'semantic': ['ac', 'ac-sign'], 'relation': '#5002->5003'}), ('é', {'lemma': 'ser', 'morpho': ['V', 'PR', '3S', 'IND', 'VFIN'], 'POS': 'FS-STA', 'semantic': ['vK', 'fmc', 'mv'], 'relation': '#5003->0'}), ('elevadíssimo', {'lemma': 'elevadíssimo', 'morpho': ['ADJ', 'M', 'S'], 'POS': '<SC', 'semantic': ['nh', 'SC'], 'relation': '#5004->5003'}), ('.', {'lemma': '.', 'morpho': ['punct'], 'POS': 'punct', 'semantic': [], 'relation': '#5005->0'})), {'1': 0, '2': 1, '1001': 2, '1002': 3, '1003': 4, '1004': 5, '1005': 6, '1006': 7, '1007': 8, '1008': 9, '2001': 10, '2002': 11, '2003': 12, '2004': 13, '2005': 14, '2006': 15, '2007': 16, '2008': 17, '2009': 18, '2010': 19, '3001': 20, '3002': 21, '3003': 22, '3004': 23, '3005': 24, '3006': 25, '3007': 26, '3008': 27, '3009': 28, '3010': 29, '3011': 30, '3012': 31, '3013': 32, '3014': 33, '3015': 34, '3016': 35, '3017': 36, '4001': 37, '4002': 38, '4003': 39, '4004': 40, '4005': 41, '4006': 42, '4007': 43, '4008': 44, '4009': 45, '4010': 46, '5001': 47, '5002': 48, '5003': 49, '5004': 50, '5005': 51})

# EXEMPLO DE SAÍDA DO PALAVRAS (BICK, 2000)

PERIGO! Dormir mais de 10h eleva risco de AVC. O AVC pode matar alguém que dorme com você. **Compartilhe** esta mensagem **urgentemente** com quem tem o hábito de dormir mais de 10 horas por dia. **Você** pode salvar a vida de alguém que ama. Este risco é elevadíssimo.

('urgentemente', {'lemma': 'urgentemente', 'morpho': ['ADV'], 'POS': '<ADVL', 'semantic': ['nh', 'ADVL'], 'relation': '#3004->3001'})

# EXEMPLO DE SAÍDA DO PALAVRAS (BICK, 2000)

PERIGO! Dormir mais de 10h eleva risco de AVC. O AVC pode matar alguém que dorme com você. **Compartilhe** esta mensagem **urgentemente** com quem tem o hábito de dormir mais de 10 horas por dia. **Você** pode salvar a vida de alguém que ama. Este risco é elevadíssimo.

('urgentemente', {'lemma': 'urgentemente', 'morpho': ['ADV'], 'POS': '<ADVL', 'semantic': ['nh', 'ADVL'], 'relation': '#3004->3001'})

('Compartilhe', {'lemma': 'compartilhar', 'morpho': ['V'], 'PR', '1/3S', 'IMP' 'VFIN'], 'POS': 'FS-STA', 'semantic': ['\*', 'vH', 'fmc', 'mv'], 'relation': '#3001->0'})

# EXEMPLO DE SAÍDA DO PALAVRAS (BICK, 2000)

PERIGO! Dormir mais de 10h eleva risco de AVC. O AVC pode matar alguém que dorme com você. **Compartilhe** esta mensagem **urgentemente** com quem tem o hábito de dormir mais de 10 horas por dia. **Você** pode salvar a vida de alguém que ama. Este risco é elevadíssimo.

('urgentemente', {'lemma': 'urgentemente', 'morpho': ['ADV'], 'POS': '<ADVL', 'semantic': ['nh', 'ADVL'], 'relation': '#3004->3001'})

('Compartilhe', {'lemma': 'compartilhar', 'morpho': ['V'], 'PR', '1/3S', 'IMP' 'VFIN'], 'POS': 'FS-STA', 'semantic': ['\*', 'vH', 'fmc', 'mv'], 'relation': '#3001->0'})

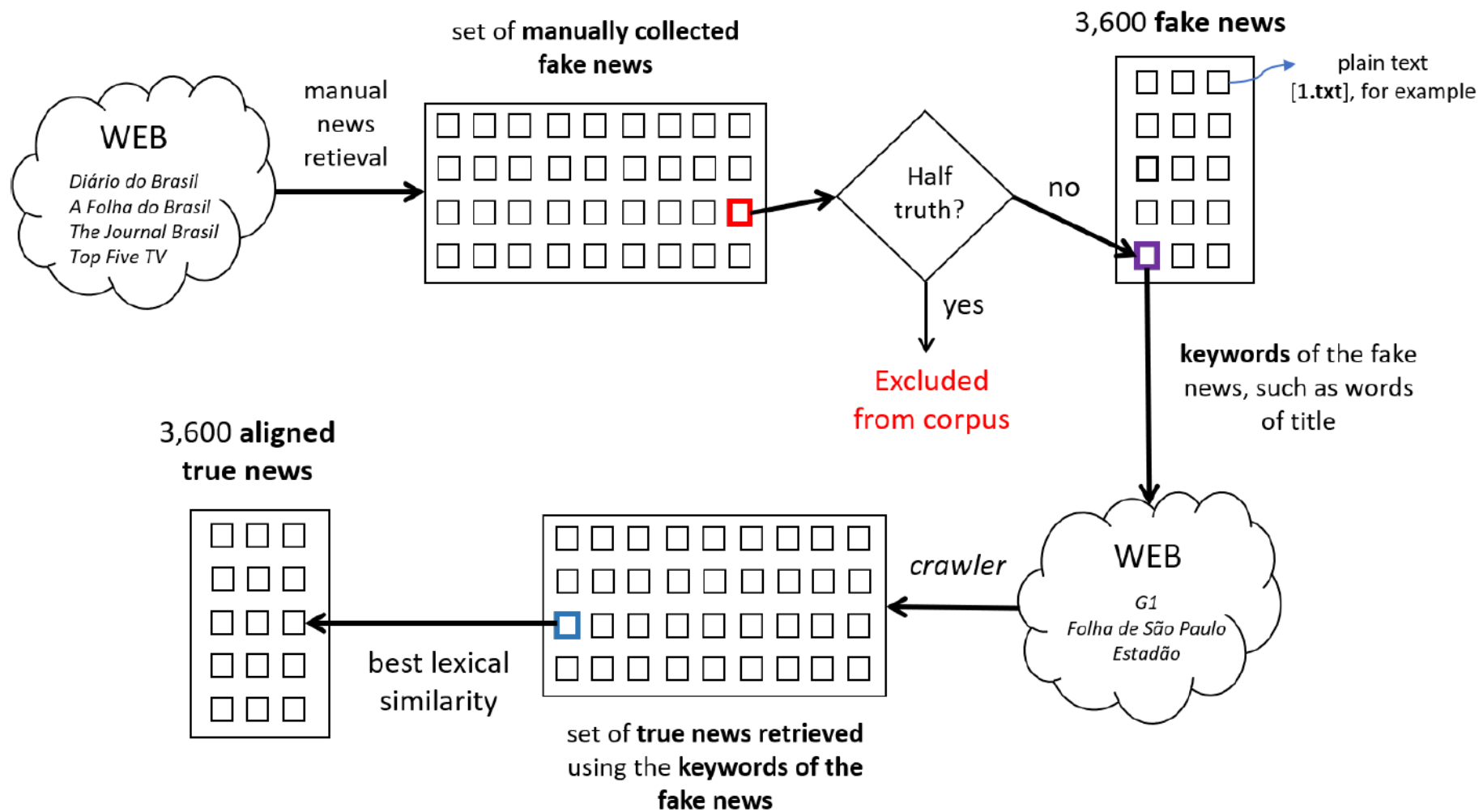
('Você', {'lemma': 'você', 'morpho': ['PERS'] 'M/F', '2S', 'NOM'], 'POS': 'SUBJ>', 'semantic': ['\*'], 'relation': '#4001->4002'})



# PRIMEIROS EXPERIMENTOS

## Fake.Br Corpus

Dataset com 7.200 notícias: 3.600 falsas e 3.600 verdadeiras



# PRIMEIROS EXPERIMENTOS

## Experimentos com vários desses atributos

Conseguimos chegar a 89% de acurácia quando usamos abordagens de aprendizado de máquina mais comuns e **97% de acurácia** quando combinamos várias abordagens.

|  | LBR          | FCR          | FPR          | F-measure    |
|--|--------------|--------------|--------------|--------------|
| Ling. features   | 0.060        | 0.941        | 0.940        | 0.941        |
| BoW (trunc. text)  | 0.057        | 0.932        | 0.943        | 0.937        |
| Ensemble – BoW (trunc. text) + ling. features              | <b>0.024</b> | 0.954        | <b>0.976</b> | <b>0.965</b> |
| Stacking – BoW (trunc. text) + ling. features              | 0.030        | <b>0.959</b> | 0.970        | 0.964        |
| Monteiro et al. (2018) (Ling. features)                    | –            | 0.53         | 0.57         | 0.55         |
| Monteiro et al. (2018) (BoW)                               | –            | 0.89         | 0.88         | 0.88         |
| Monteiro et al. (2018) (POS tags + semantic classes + BoW) | –            | 0.89         | 0.88         | 0.89         |

# O FakeCheck – Detector de Fake News



## Detector de Fake News

### Como funciona?

**Copie o texto de uma notícia, cole na caixa abaixo e clique em "Enviar".** O sistema irá processar o texto para identificar características de escrita, como palavras usadas ou classes gramaticais mais frequentes, e utilizar essas características em um modelo de aprendizado de máquina que classificará a notícia em verdadeira ou falsa. Para mais informações sobre como o sistema funciona e sua taxa de acerto, clique [aqui](#). Você também pode utilizar o nosso [bot do WhatsApp](#).

**ATENÇÃO: Utilize o texto completo da notícia! O texto deve ter pelo menos 100 palavras.** O sistema pode não funcionar corretamente com apenas partes de notícias.

Insira o texto da sua notícia aqui.

Notícia

Modelo de Detecção

Palavras do Texto

ENVIAR ➤

### Financiamento e Apoio



# ABRE PARÊNTESES...

Intercâmbio na Universidade de Sheffield, Reino Unido



## Nutrition Labels

# ABRE PARÊNTESES...

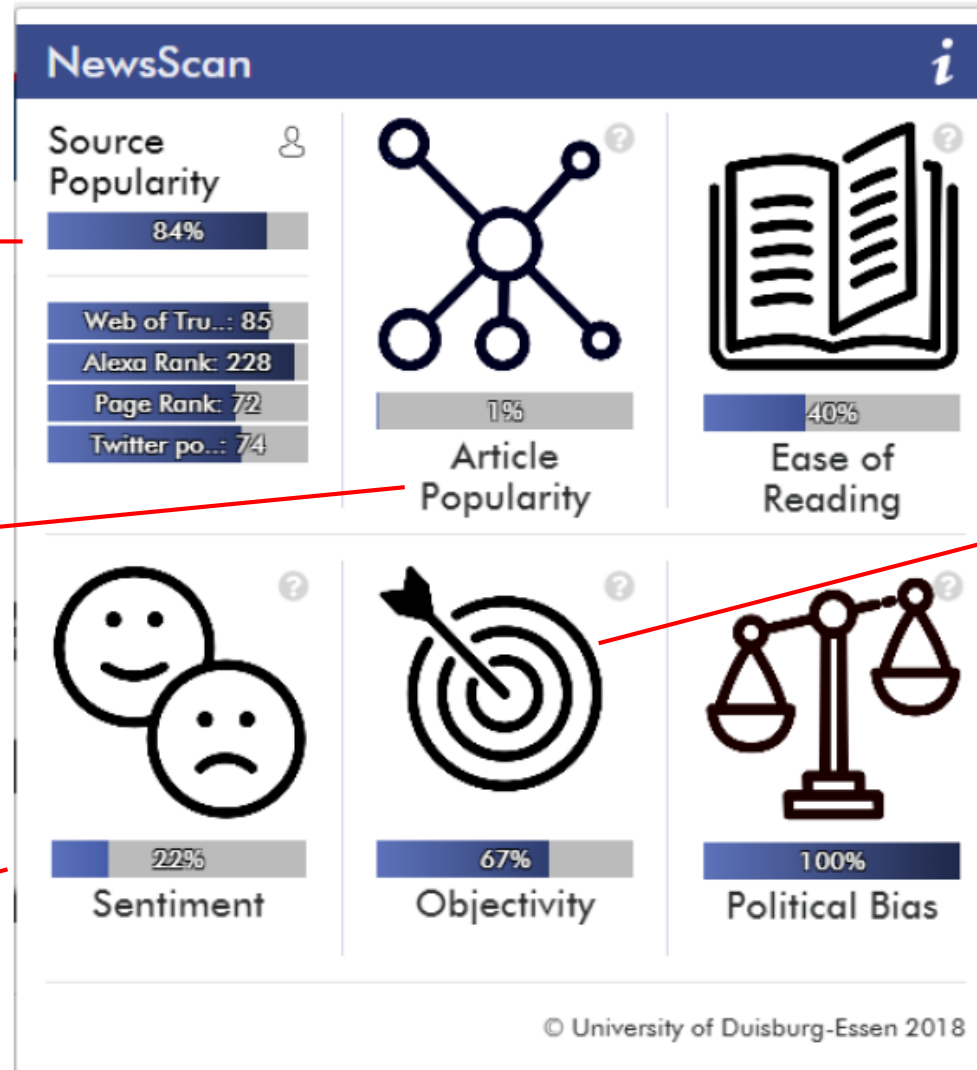
Intercâmbio na Universidade de Sheffield, Reino Unido



Reputação e influência da origem da notícia

Popularidade da notícia nas redes sociais (tweets por hora, por exemplo)

Notícia é escrita com palavras que indicam positividade ou negatividade



Facilidade de leitura e entendimento da notícia

A notícia é escrita de forma neutra, sem expressões como “em minha opinião” ou “eu acho”.

A notícia tem um lado político iminente (esquerda ou direita)

# ABRE PARÊNTESES...

Intercâmbio na Universidade de Sheffield, Reino Unido



Análise da **facilidade de leitura de uma notícia.**

Uma notícia verdadeira tem que **ser mais fácil ou mais difícil** de ler que uma notícia falsa?

# BASEADA EM LINGUÍSTICA – TENTATIVA 2

## Legibilidade em Notícias Falsas



- Qual o impacto dos atributos de **legibilidade** na detecção de notícias falsas?
- Utilização de medidas clássicas:
  - Flesch Ease Reading (Flesch, 1979)
  - Brunet's Index (Brunet, 1978)
  - Honoré Statistic (Honoré, 1979)
  - Dale Chall Formula (Dale and Chall, 1948)
  - Gunning Fog Index (Gunning, 1968)
- Medidas de Coesão do texto e Psicolinguísticas
  - Psicolinguísticas: medição do processamento cognitivo humano
    - Concretude, Imageabilidade, Familiaridade e Idade de Aquisição de cada palavra
- Extraídas por meio do **NILC-Metrix**

# BASEADA EM LINGUÍSTICA – TENTATIVA 2

## Legibilidade em Notícias Falsas



Todos os 165 atributos  
do NILC-Metrix

| Experiment                               | Precision   |             | Recall      |             | F-Measure   |             | Accuracy    |
|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|  | Fake        | True        | Fake        | True        | Fake        | True        |             |
| Monteiro et al. (2018)                   | 0.89        | 0.89        | 0.89        | 0.89        | 0.89        | 0.89        | 0.89        |
| E01                                      | 0.94        | 0.89        | 0.88        | 0.94        | 0.91        | 0.91        | 0.91        |
| E03                                      | 0.92        | 0.92        | 0.92        | 0.92        | 0.92        | 0.92        | 0.92        |
| E04                                      | 0.93        | 0.90        | 0.90        | 0.93        | 0.91        | 0.92        | 0.92        |
| <b>E05: Monteiro et al. (2018) + E01</b> | <b>0.95</b> | <b>0.91</b> | <b>0.91</b> | <b>0.95</b> | <b>0.93</b> | <b>0.93</b> | <b>0.93</b> |
| <b>E06: Monteiro et al. (2018) + E03</b> | <b>0.93</b> | <b>0.91</b> | <b>0.91</b> | <b>0.93</b> | <b>0.93</b> | <b>0.93</b> | <b>0.93</b> |
| E07: Monteiro et al. (2018) + E04        | 0.89        | 0.90        | 0.91        | 0.88        | 0.90        | 0.89        | 0.89        |

23 atributos  
selecionados por  
abordagens de  
*feature selection*

17 atributos de  
legibilidade selecionados  
do NILC-Metrix: clássicos,  
coesão e psicolinguísticos



# NOVAS ABORDAGENS

Baseada em conteúdo

Precisamos verificar além do texto: conhecimento de mundo

# NOVAS TENTATIVAS

## Baseada em conteúdo

Precisamos verificar além do texto: conhecimento de mundo

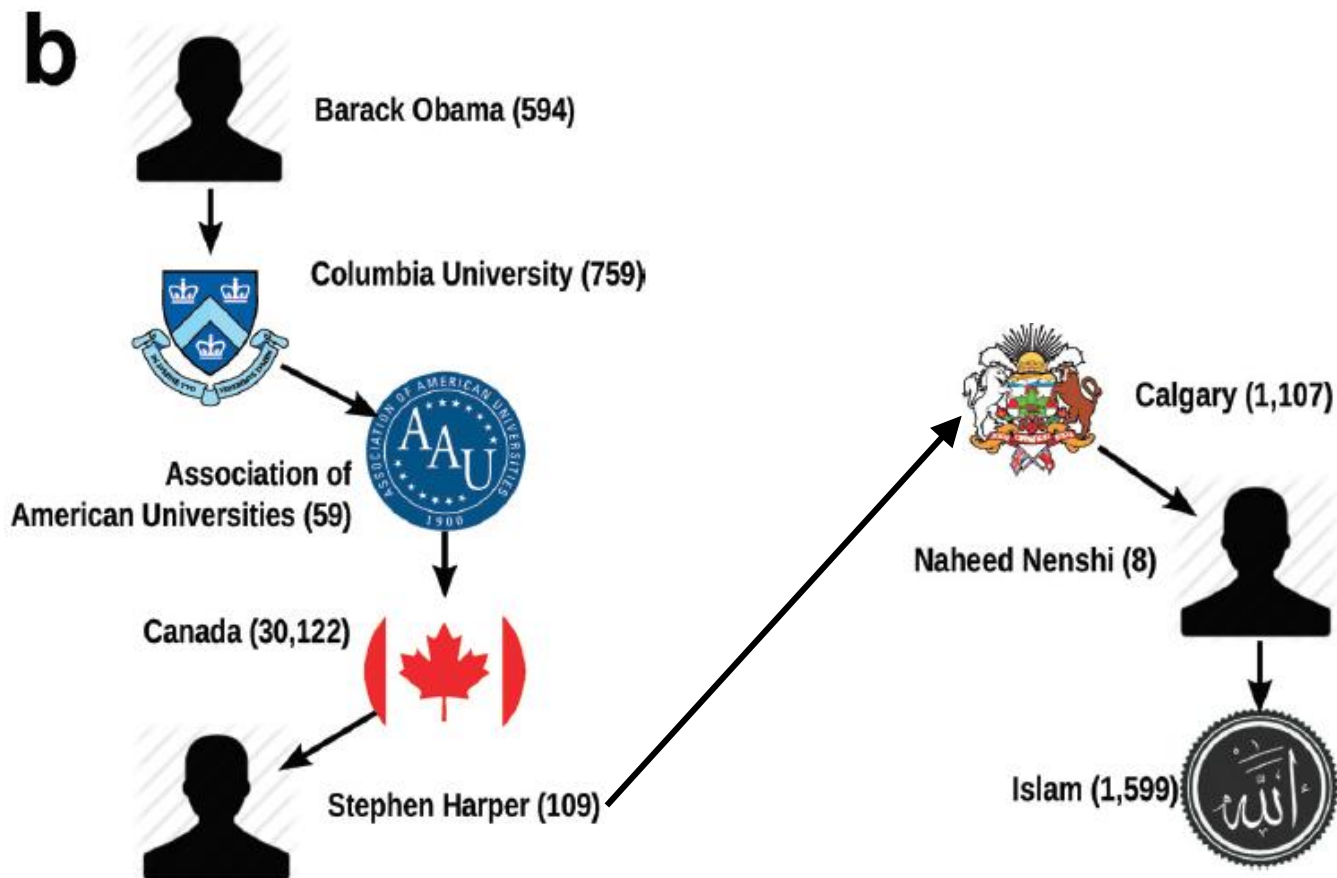
## Grafos de Conhecimento

Todos os eventos presentes nas notícias são mapeados em um grafo

No teste, quando uma nova notícia é submetida ao grafo, verifica-se a existência dos eventos. Caso não exista nenhum indício do evento, a notícia **pode ser** falsa ou mesmo recente.

Princípio: **tripla (sujeito, relação, objeto)**

# GRAFOS DE CONHECIMENTO



O valor de verdade mede a proximidade semântica da tripla a partir da distância que os argumentos estão no caminho mais curto entre eles no grafo, levando em conta o grau de cada nó participante do caminho.

# NOVAS ABORDAGENS

## Baseada em conteúdo

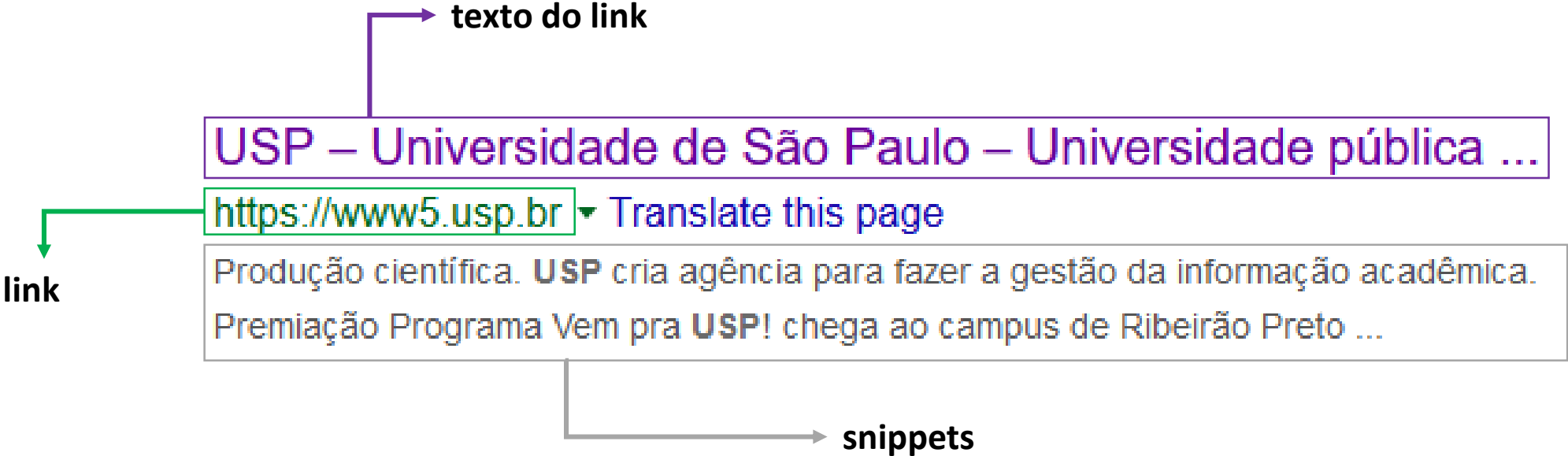
Precisamos verificar além do texto: conhecimento de mundo

## Pesquisa no Google

Reprodução do comportamento usual das pessoas quando usam o Google para verificar a veracidade da notícia

Utilização dos *snippets* para análise: caso haja alguma palavra que indique que a notícia é falsa, ela é classificada como falsa.

# PESQUISA NO GOOGLE



# NOVAS ABORDAGENS

## Baseada em conteúdo

Precisamos verificar além do texto: **conhecimento de mundo**

|   | Experiment    | P    | R    | F1   | A    |             |
|---|---------------|------|------|------|------|-------------|
| Título completo da notícia  | E1            | Fake | 0.71 | 0.12 | 0.20 | 0.53        |
|   |               | True | 0.52 | 0.95 | 0.67 |             |
| Título da notícia sem <i>stopwords</i>  | E2            | Fake | 0.71 | 0.13 | 0.23 | 0.54        |
|   |               | True | 0.52 | 0.94 | 0.67 |             |
| Título da notícia sem <i>stopwords</i> e se tinham ocorrências em sites de <i>fact-checking</i> | <b>E2-alt</b> | Fake | 0.76 | 0.24 | 0.36 | <b>0.68</b> |
|   |               | True | 0.60 | 0.98 | 0.74 |             |
| Título completo na ordem exata  | E3            | Fake | 1.00 | 0.01 | 0.01 | 0.51        |
|   |               | True | 1.00 | 1.00 | 1.00 |             |

Resultados - Pesquisa no Google

# NOVAS ABORDAGENS

## Grafos de Conhecimento

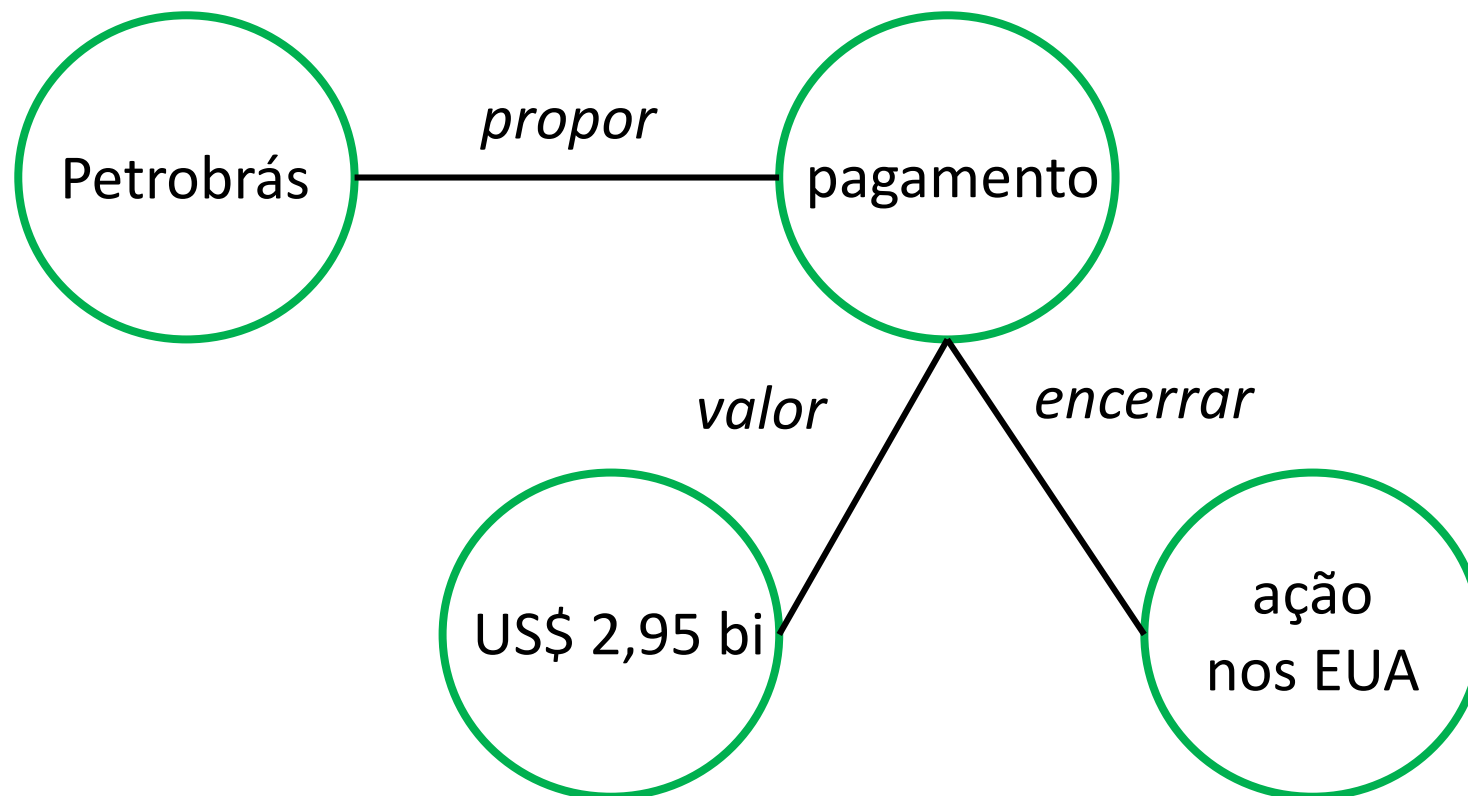
Em andamento/finalizando. Primeiros experimentos foram com afirmações. Pretendemos utilizar **notícias de um período de tempo** para criar o grafo de conhecimento e analisar sentença por sentença e, no final, o sistema responder com uma **sugestão – auxílio ao usuário**.

## Pesquisa no Google

O baixo valor de acurácia (comparado aos outros métodos) mostra que precisamos **analisar/ler/pesquisar o que nos é mostrado** pelas ferramentas de pesquisa!

# NOVAS ABORDAGENS – GRAFO DE CONHECIMENTO

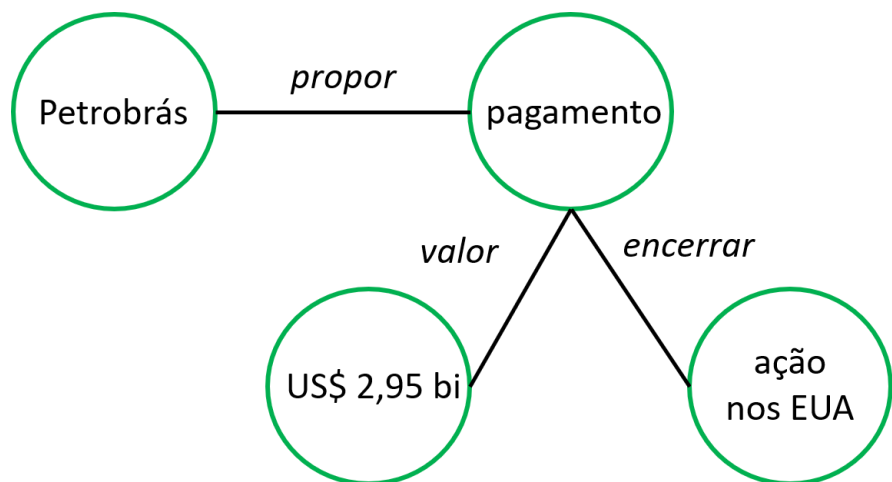
Petrobrás propõe pagamento de US\$ 2,95 bi para encerrar ação nos EUA





# NOVAS ABORDAGENS – GRAFO DE CONHECIMENTO

Petrobrás propõe pagamento de US\$ 2,95 bi para encerrar ação nos EUA



Estrutura de dados para a formação do grafo: **dicionário**

```
{ 'Petrobrás' : [[ 'pagamento' , 'propor' ] ,  
'pagamento' : [[ 'Petrobrás' , 'propor' ] ,  
[ 'US$ 2,95 bi' , 'valor' ] , [ 'ação nos EUA' ,  
'encerrar' ] ] }
```

Arestas com a relação entre os nós contém o token **lematizado**

propõe -> propor

encerrar -> encerrar

# NOVAS ABORDAGENS – GRAFO DE CONHECIMENTO

|                         | <b>Acurácia</b> |
|-------------------------|-----------------|
| <b>Fake.Br Corpus</b>   | <b>0.68</b>     |
| <b><i>Categoria</i></b> |                 |
| Política                | 0.64            |
| TV e Celebidades        | 0.76            |
| Sociedade e Cotidiano   | 0.67            |
| Ciência e Tecnologia    | 0.72            |
| Economia                | 0.77            |
| Religião                | 0.77            |

Resultados – Reprodução de Ciampaglia (2015), com a utilização das notícias do Fake.Br.

Triplas extraídas por meio da abordagem *DptOIE*, que utiliza noções de Open Information Extraction.

# NOVAS ABORDAGENS – PRÓXIMOS PASSOS

## Grafos de Conhecimento

Porém, percebe-se que as informações das arestas (relações) **não são levadas em consideração...**

Além desse método, que está em construção, ainda temos em andamento/finalizando:

- Substituição dos sujeitos/objetos por seus vetores de embeddings (**word2vec** ou **BERT**).
  - Experimentos chegam a **75% de acurácia** no  $t(true) > t(fake)$
- Utilização de noções de **Redes Complexas**.
  - Primeiros testes alcançaram **69% de acurácia** em notícias **truncadas**

# MAS TEMOS OUTRAS QUESTÕES...

- **Limites do Processamento de Linguagem Natural (PLN) para a tarefa...**
  - A mentira está só (mais?) na língua
    - E o olhar, o gesto, etc.? Quem cuida?
- **Limites do humano para a tarefa!**
  - Humano com 54% de acerto, com variação do domínio
- **Como tratar**
  - ~~Meias verdades,~~
  - Ironia, sarcasmo, sátira, paródia
  - Revisões falsas,
  - Tempo das notícias,
  - Pós-verdades...?
- **Aplicativo para indicar mentira vs verdade (WhatsApp)**
  - Isso é censura? É ético indicar mentira o tempo todo?

# PARA ISSO, TEMOS AJUDA!

- Grupo multidisciplinar USP + UFSCar
- A tentativa de resolução desse problema faz parte de dois grandes projetos: **OPINANDO** e **C4AI**
- **E estamos aberto a colaborações!**

# MUITO OBRIGADO!

Perguntas, dúvidas, sugestões, críticas...! Espaço aberto!

## Links úteis:

- FakeCheck: <http://nilc-fakenews.herokuapp.com/>
- Nosso primeiro trabalho publicado:  
<http://conteudo.icmc.usp.br/pessoas/taspardo/PROPOR2018-MonteiroEtAl.pdf>
- O Fake.Br Corpus: <https://github.com/roneysco/Fake.br-Corpus>