

Professor(a):

José Wellington Franco da Silva

wellington@crateus.ufc.br



Análise de Dados Textuais

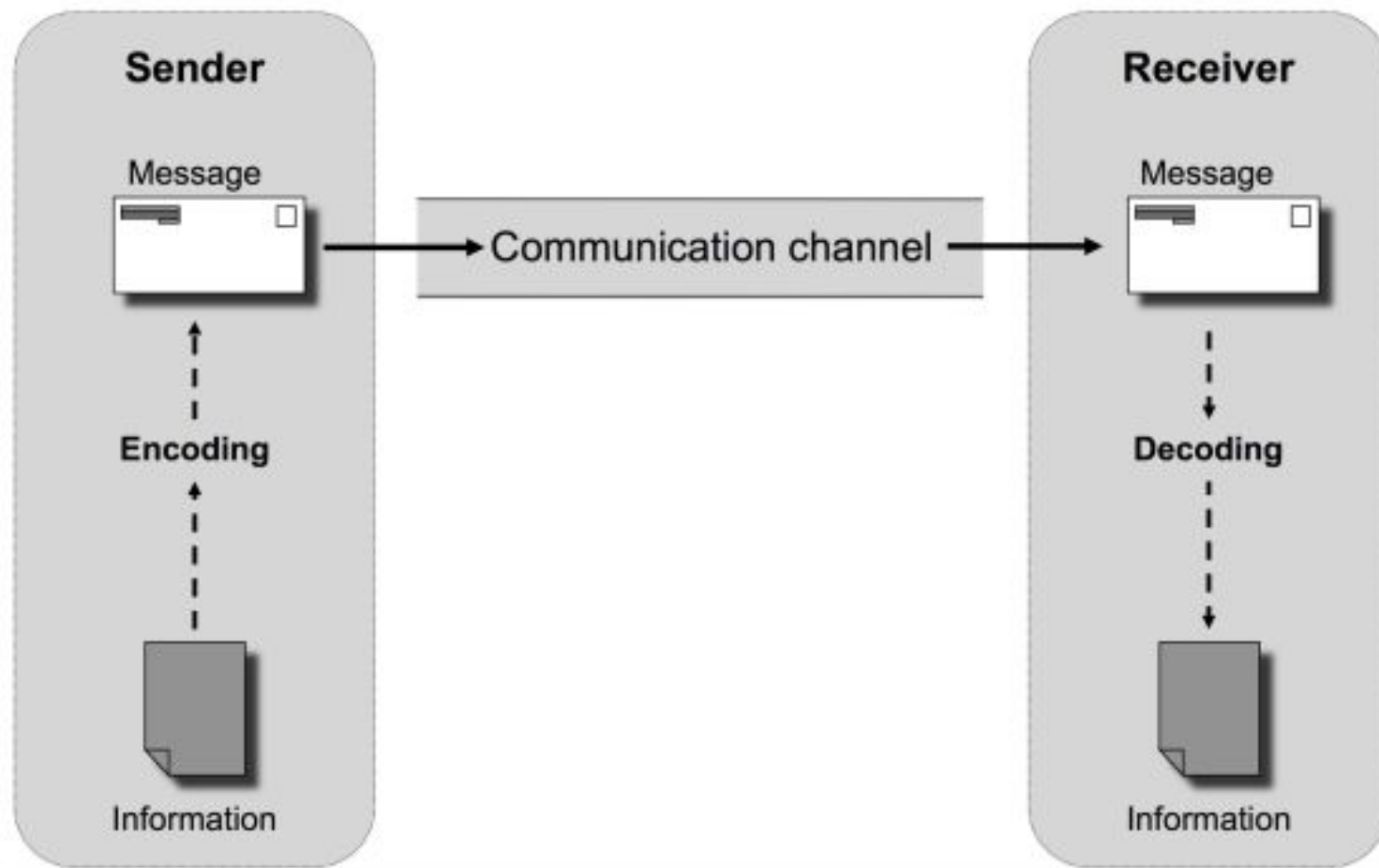
ANO:2019

O que é a Comunicação?

- A comunicação é um processo pelo qual as **informações** são **trocadas** entre indivíduos através de um **sistema comum de símbolos, sinais ou comportamento**. [Merriam - Webster]
- Comunicação é o ato de **transmitir significados pretendidos** de um entidade ou grupo para outro através do **uso de sinais e regras semióticas compreendidos mutuamente**. [wikipedia]



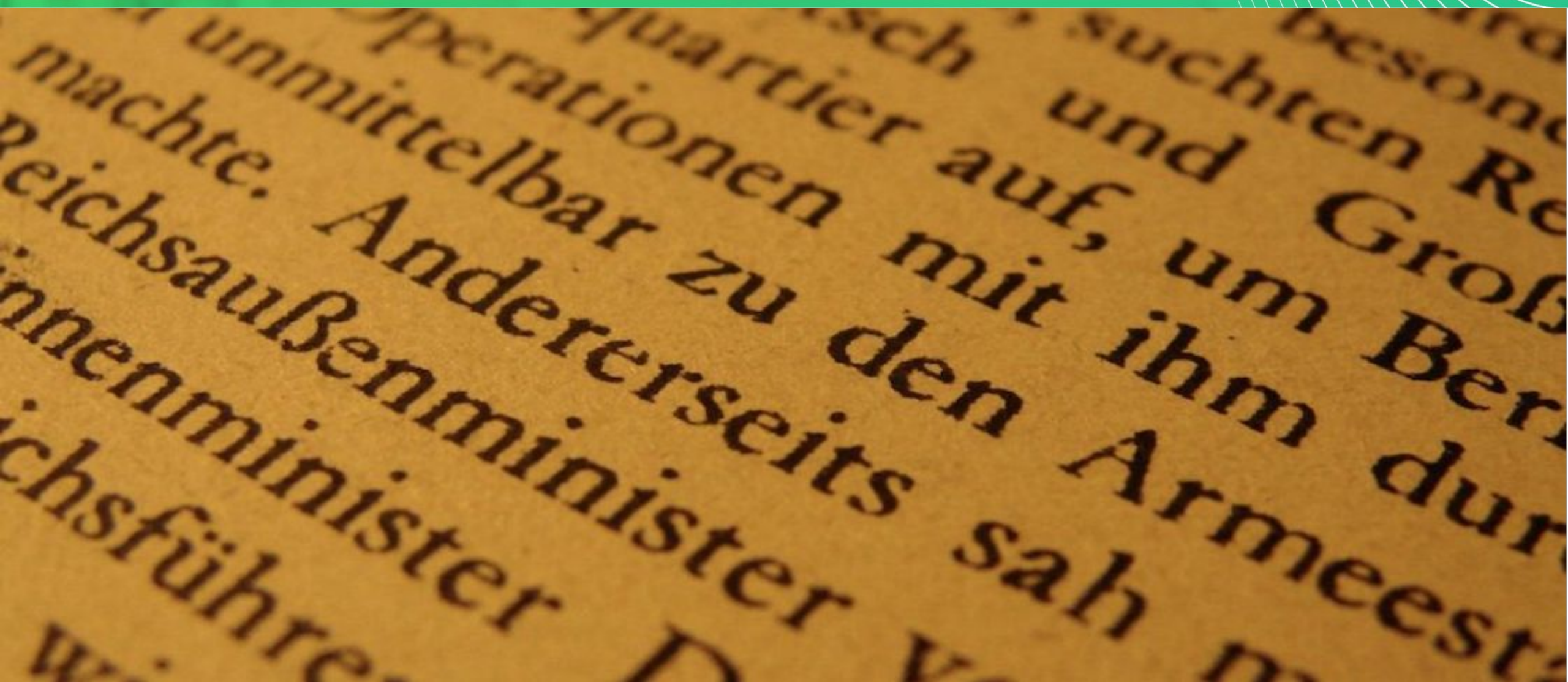
Processo de comunicação



Ch. Meinel, H. Sack:
Digital Communication -
Communication Multimedia,
Security, Springer, 2014.

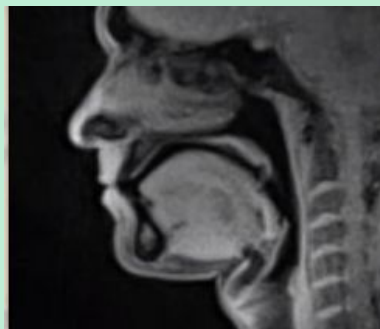
- [illegible]

Linguagem Natural (LN)



- **Fala**

- **Escutando**
- **Falando**



- **Texto**

- **Lendo**
- **Escrevendo**



Como é codificada a informação textual?

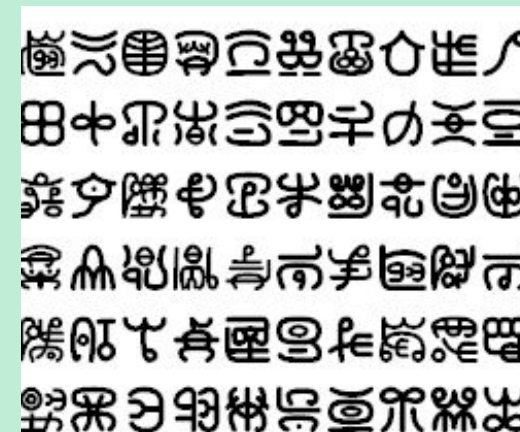
- **Sistemas de escrita alfabética**

- **Caractere único refere-se a um único som (alfabeto fonêmico);**
- **Alfabetos consoantes, vogais podem ser deduzidas do contexto;**



Como é codificada a informação textual?

- **Sistema de Escrita Silábica e Logográfica**
 - Os sistemas silábicos também envolvem um mapeamento entre caracteres e som;
 - Um logograma é um símbolo que representa uma unidade de significado.



O que é a informação?

- **Informação** é aquela que informa;
- As **informações** são transmitidas como conteúdo de uma mensagem;
- As **informações** podem ser codificadas em várias formas para transmissão e interpretação;
- A **informação** é qualquer propagação de causa e efeito dentro de um sistema;
- A informação reduz incerteza.

O que significa entender?

- **Um correto entendimento depende de:**
 - **Sintaxe;**
 - **Semântica;**
 - **Contexto;**
 - **Pragmatismo;**
 - **Experiência**

- Na **gramática**
 - A sintaxe denota o estudo dos princípios e processos pelos quais as frases são construídas em idiomas específicos;
- Em **linguagens formais**
 - A sintaxe é apenas um conjunto de regras, pelas quais expressões bem formadas podem ser criadas a partir de um conjunto fundamental de símbolos (alfabeto);
- Na **computação**
 - A sintaxe define a estrutura normativa dos dados.

- Faz parte da linguística focada no sentido e no significado da linguagem ou nos símbolos da linguagem;
- É o estudo da interpretação de sinais ou símbolos usado por agentes ou comunidades em circunstâncias e contextos particulares;
- A **semântica** pergunta como o sentido e o significado de conceitos complexos podem ser derivados de conceitos simples baseados nas regras de sintaxe.
- A **semântica** de uma **mensagem** depende do contexto e da pragmática.

- Denota o entorno de um símbolo em uma expressão respectiva relacionamento com expressões circundantes e outros elementos relacionados;
- O contexto denota todos os elementos de qualquer tipo de comunicação que definem a interpretação do conteúdo comunicado.

- Reflete a intenção pela qual o idioma é usado para comunicar uma mensagem;
- Em linguística, pragmática denota o estudo da aplicação da linguagem em diferentes situações;
- Também denota o objetivo do orador;
- A pragmática estuda as maneiras pelas quais o contexto contribui para o significado.

- **A experiência considera todas as informações que você aprendeu e colocou em contexto com o mundo em que vive;**
- **A experiência nesse sentido também costuma ser chamada de conhecimento do senso comum ou conhecimento do mundo.**

- **Para uma comunicação bem-sucedida**
 - **As informações devem ser transmitidas corretamente (sintaxe);**
 - **O significado (semântica) da informação transmitida deve ser interpretado corretamente (= entendimento);**
- **O entendimento**
 - **O contexto do remetente e do destinatário;**
 - **A pragmática do remetente;**

- **Ruslan Mitkov, The Oxford Handbook Of Computational Linguistics, Oxford University Press, 2003.**
- **Robert Dale, Hermani Moisi, Harold Somers, Handbook Of Natural Language Processing, Markcel Dekker Inc.**
- **James Allen, Natural Language Processing, Pearson Education, 2003.**
- **Christopher D.Manning & Henrich Schutze, Foundations Of Statistical Natural Language Processing, The MIT Press, 2001**
- **Douglas Biber, Susan Conrad, Randi Reppen, Corpus Linguistics – Investigating Language Structure And Use, Cambridge University Press, 2000.**

- **David Singleton, Language And The Lexicon: An Introduction, Arnold Publishers, 2000.**
- **Andrew Radford, Minimalist Syntax: Exploring The Structure Of English, Cambridge University Press, 2004.**
- **Rebecca Stott & Peter Chapman, Grammar And Writing, Pearson Education, 2001.**
- **A. Athithan, Linguistic Structures In Tamil – A Historical Study, Madurai Kamaraj University, 1989.**
- **R.Kothandaraman, Tamil Syntax: New Perspectives, Pondicherry Institute Of Linguistics And Culture, 1990.**

- **Raymond Hickey, Corpus Presenter: Software for language analysis with a manual and a Corpus of Irish English as sample data, John Benjamins Publishing Company, 2003.**
- **Guy Aston & Lou Burnard, The BNC Handbook: Exploring the British National Corpus with SARA, Edinburgh University Press, 1998.**
- **Donny D. Steinberg, Hiroshi Nagata, David P. Aline, Psycholinguistics: Language, Mind and World, Second Edition, Pearson Education, 2001.**
- **Cyruk Goutte, Nicola Cancedda, Marc Dymetman, and George Foster, Learning Machine Translation, PHI, 2010.**

Dúvidas são bem vindas!!

- wellington@crateus.ufc.br

