Uso de linguística de corpus no aprendizado da escrita científica em inglês

Osvaldo N. Oliveira Jr.

Núcleo Interinstitucional de Lingüística Computacional

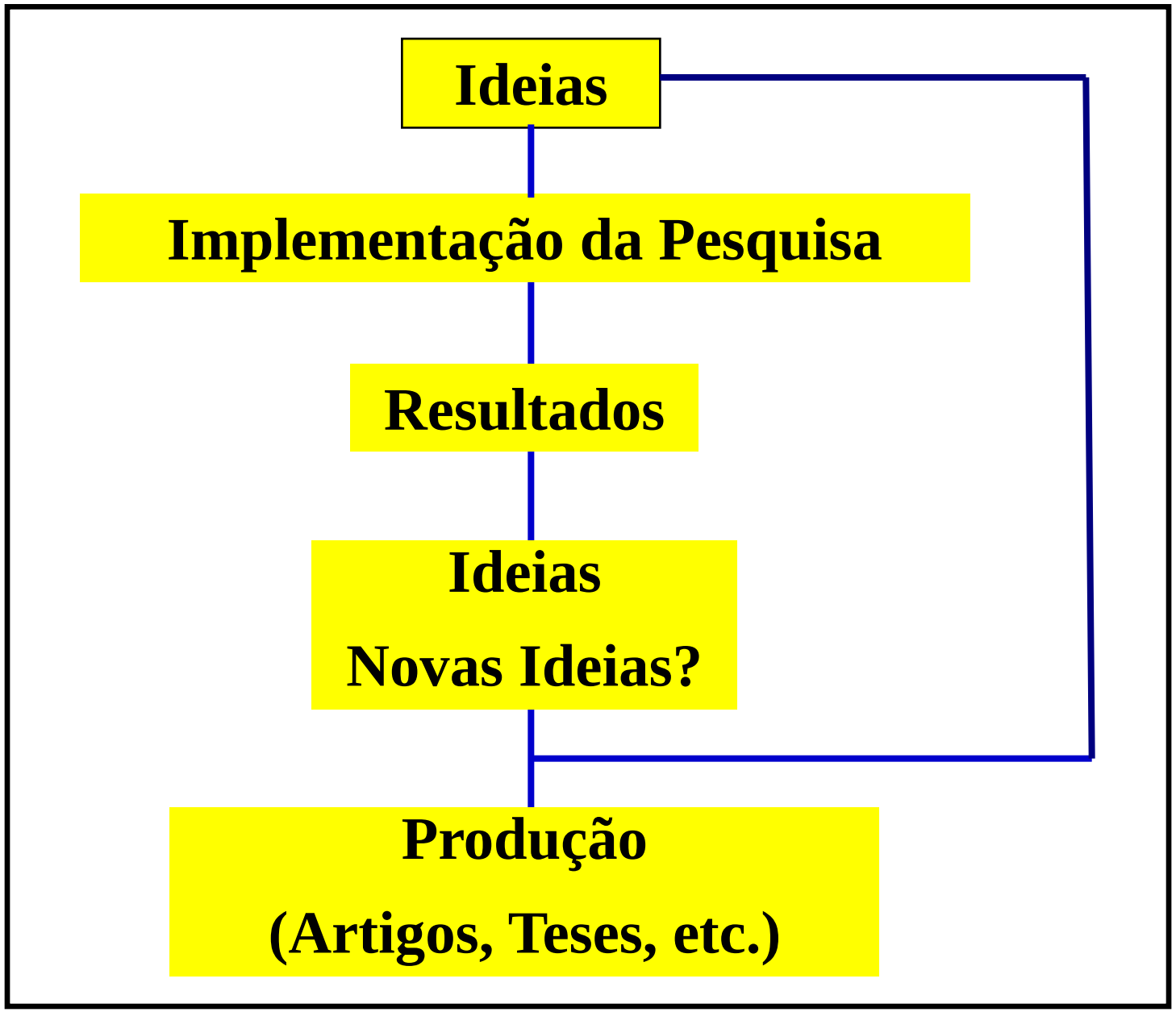
USP-São Carlos, UFSCar, Unesp-Araraquara

http://nilc.icmc.usp.br

Escolha do problema científico

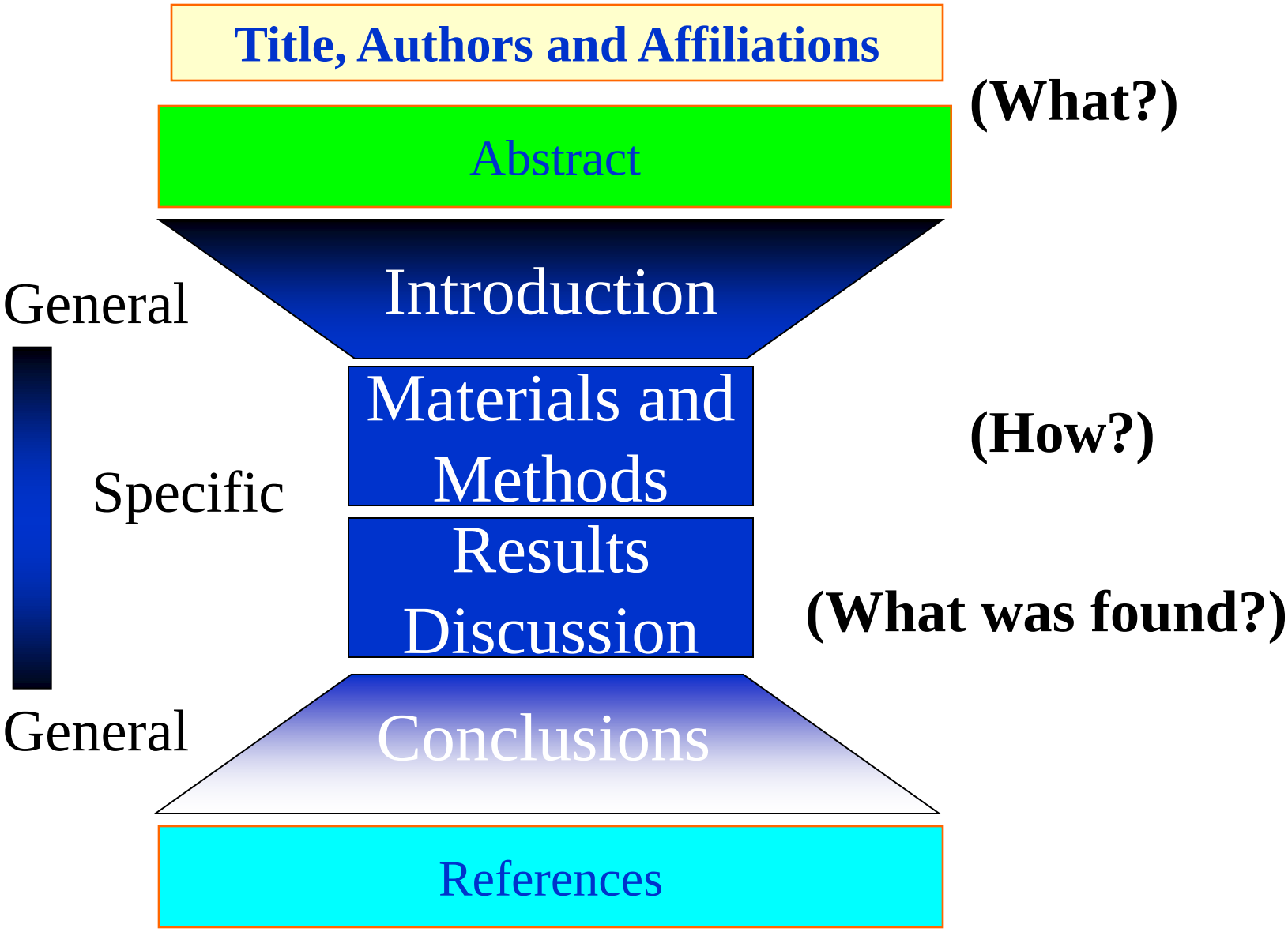
- **O que quero provar?**
- **Quais são os objetivos genéricos?**
- **Que métodos empregarei?**
- **Se bem sucedido, onde publicar?**

Publicando ideias, conceitos



Cientistas publicam ideias e conceitos; NÃO resultados!

Organização de um artigo



Hill et al.,, TESOL Quarterly, 16: 333, 1982:

Cronograma da elaboração

- **“brain-storming”**
- **coleta e descrição dos resultados**
- **resumo do que já se obteve na literatura**
- **elaboração da estrutura do artigo**

Componentes do artigo

- **Resumo (fazer no final)**
- **Introdução (fazer no final)**
- **Fundamentação teórica, revisão da literatura, tópicos introdutórios**
- **Metodologia**
- **Resultados e discussão**
- **Conclusões**

O Resumo

- **Informativo X descritivo**
- **Deve conter todas as conclusões principais**
- **Equilíbrio entre detalhes fornecidos de resultados, metodologia, etc.**

O Resumo

- *Contextualização*
- *Objetivo*
- *Metodologia*
- *Resultados e contribuições*
- *Conclusão*

Resultados e Discussão devem ser organizados de acordo com sequência lógica

Observem que:

Sequência Lógica	Nem sempre =	Sequência Cronológica

A Introdução

- **Funções distintas**

  artigo, ensaio, dissertação, capítulo, etc.

- **Contextualização**

- **Revisão da literatura**

- **Descrição dos objetivos**

Fundamentação

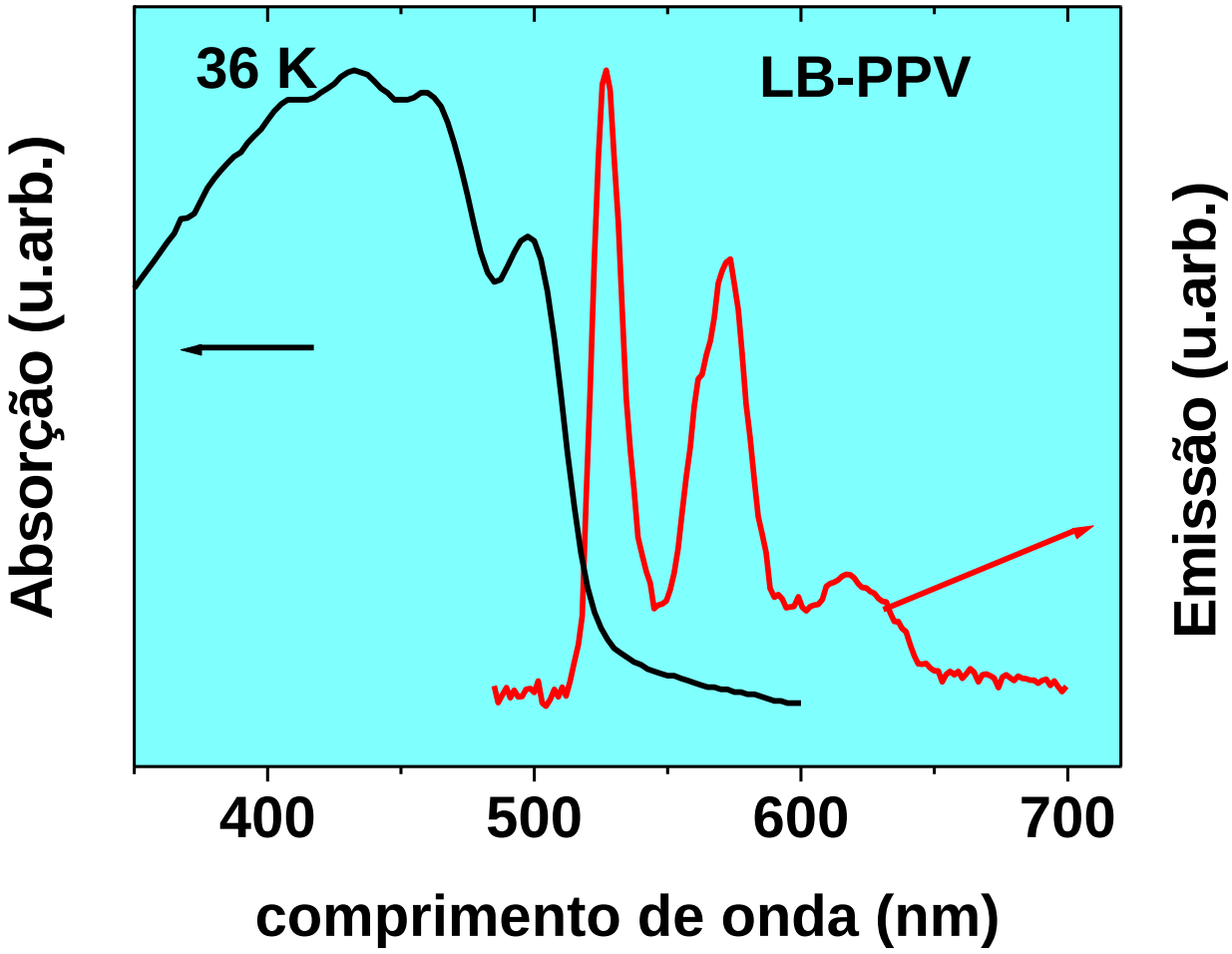
Fornecimento de informações importantes para acompanhar os resultados

- **Detalhamento de teorias**
- **Revisão da literatura**

Metodologia e Resultados

- **Na primeira edição - descrição simples**
- **Descreva cada figura, tabela ...**
- **Analise cada resultado**
- **Confronte com a literatura**
- **Discuta com base nas hipóteses**

Descrição de um Resultado



Discussão e Conclusões

- **Verificar consistência dos argumentos**
- **Comentar concordância ou discrepâncias com a literatura**
- **Concluir só o que já foi mostrado**

O Texto

- **Verificar problemas de superfície**
- **Paragrafação**
- **Coesão e coerência**
- **Evitar**

  zigue-zague, repetição, sentenças longas, palavras desnecessárias

Edição do Texto

- **Verificar problemas de superfície**

  zigue-zague, repetição, sentenças longas, palavras desnecessárias

- **Formatação**

- **Encadeamento de idéias**

- **Verificar estrutura do artigo**

Estereotipagem dos Procedimentos

- **Desejável ou necessário?**
- **Vencendo o bloqueio inicial**
- **Garantia de consistência**
- **Escrita técnica - não é arte!**

Facts

- Academics must write in English – *lingua franca* of Science.
- The need to write scientific papers in English represents a considerable barrier for non-native English authors.

The main difficulties faced by Brazilian writers

- Lexical
  - Misuse of false cognates and homophone words; lack or misuse of idioms and other collocations employed in scientific texts.
- Syntactic
  - Use of grammatical constructions from mother tongue; word by word translation; over-long/over-complex sentences.
- Textual
  - Use of rhetorical structures or strategies of the mother tongue; misuse of logical relations between sentences or phases; lack of references.

Major difficulties in writing

- Surface errors (typos and grammatical errors)
- Long sentences and short, inadequate paragraphs
- Excessive number of unnecessary words
- Problems in text cohesion, inadequate use of markers and “zig-zag” in the discussion
- Lack of coherent “story” for the text

First Draft?

- Books and software tools provide help for text post-edition or hints on how to write a paper.
  - But normally fail to provide a “hands on” approach that helps the author in producing a first draft.
- To write scientific papers in English it may not be enough:
  - To be fluent in English in another text genre
  - To know the global structure of papers in the mother tongue

Corpus Linguistics - learning by example

Corpus linguistics is the study of language as expressed in samples *(corpora)* of "real world" text (from Wikipedia).

The AMADEUS strategy

Aluísio & Oliveira, 1995

(www.nilc.icmc.usp.br)

Learning by doing

USP São Carlos

Universidade Federal de São Carlos

unesp Araraquara

Learning by doing

It takes only a few minutes

to say how to do it

BUT

It may take years to learn how to do it

http://arxiv.org/abs/cs/0611013

Learning by doing

A well-written scientific paper and its abstract should follow an underlying organization or structure to convey its content. This means that there are identifiable parts in the paper describing the work. Furthermore, these parts should be organized in an ordered sequence such as:

- Introduction
    1. Problem definition
    2. Previous approaches
    3. Critique: why you/anyone still needs to work on this?
- Contribution (what had as “Gap”): how this addresses 3 above
- Methods (and materials)
- Results
- Conclusions
    1. What has been done and implications
    2. Future work

USP São Carlos
Universidade Federal de São Carlos
UFSCar
unesp Araraquara

Learning by doing

With corpus linguistics:

User guide for identifying the underlying **structure** of a well-written paper.

Focus on the **language** and its use for your own writing.

Learning by doing

1st step:

select well-written texts from reliable sources and produced by native speakers.

read the material critically, annotating expressions that convey important messages for your own work..

USP São Carlos
Universidade Federal de São Carlos
UFSCar
unesp Araraquara

Learning by doing

2nd step:

compile the expressions and sentences, clearly marking the distinct messages. This procedure should be part of your learning life – never stop doing it.

3rd step:

classify the materials according to the schemata of a scientific paper

USP São Carlos
Universidade Federal de São Carlos
UFSCar
unesp Araraquara

Classification

Strategy 1

assign the expressions to the pre-defined scheme for the various parts of an article, together with the selection (e.g. an expression taken from a component from the Introduction is automatically classified as such)

Advantage: easier and quicker
Disadvantage: user does not practice reshuffling the material

Classification

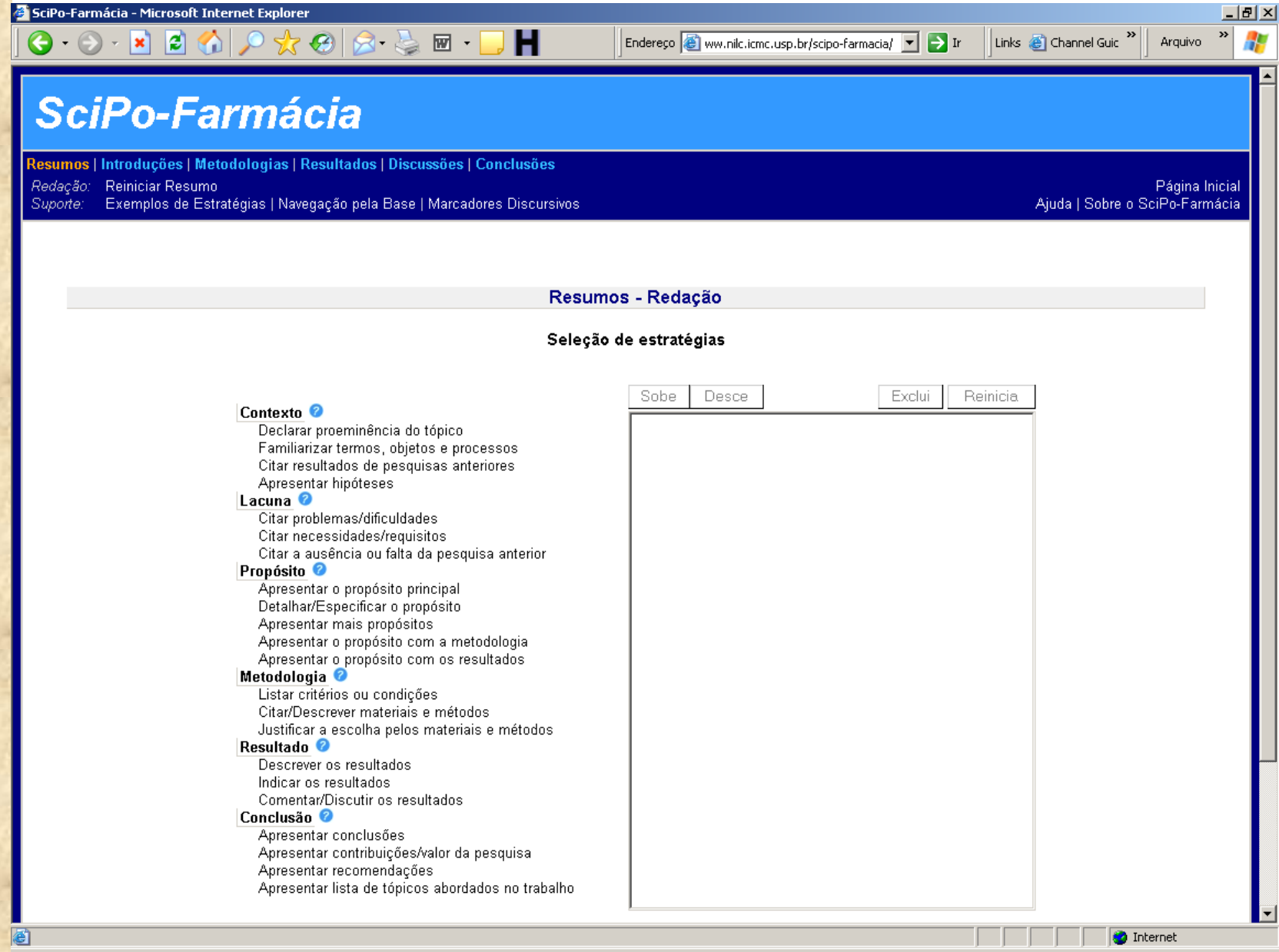
Strategy 2

select a large number of expressions (hundreds!) and only classify them later

Advantage: more efficient to learn how to interpret the messages.

Disadvantage: more time-consuming

Interface - Scipo-Farmácia



São Carlos

Universidade Federal de São Carlos

unesp Araraquara

Expressões para Introdução

a) Importance of the field, general interests, etc.
There has been substantial interest in the fabrication of ...

Contexto

b) Previous reports on related work.
Several papers have reported measurements aimed at obtaining evidence for, and insight into, ... processes in ...

Revisão da Literatura

c) What is lacking in the field.
Although significant advances have been made in the understanding of how ... (something) influences ... (another), very little further attention appears to have been given to the ...

Lacuna(s)

d) What the present work does.
The purpose of the work reported here was to study the influence of ... on the ...

Objetivo

f) Layout or Outline of the paper
The organisation (outline) of the (this) paper is as follows. In Section II we describe ... The ... is presented in section III. In Sec. II we solve the ... equation giving expressions for ... This is necessary for the work of Sec. III, in which the extended ... equation is derived. Numerical results of the theory are given in Sec. IV, together with a comparison with ... and... calculations.

Estrutura

Abstract

Self-assembly of components larger than molecules into ordered arrays is an efficient way of preparing microstructured materials with interesting mechanical and optical properties. Although crystallization of identical particles or particles of different sizes or shapes can be readily achieved, the repertoire of methods to assemble binary lattices of particles of the same sizes but with different properties is very limited. This paper describes electrostatic self-assembly of two types of macroscopic components of identical dimensions using interactions that are generated by contact electrification. The systems we have examined comprise two kinds of objects (usually spheres) made of different polymeric materials that charge with opposite electrical polarities when agitated on flat, metallic surfaces. The interplay of repulsive interactions between like-charged objects and attractive interactions between unlike-charged ones results in the self-assembly of these objects into highly ordered, closed arrays. Remarkably, some of the assemblies that form are not electroneutral—that is, they possess a net charge. We suggest that the stability of these unusual structures can be explained by accounting for the interactions between electric dipoles that the particles in the aggregates induce in their neighbors.

G.M. Whitesides et al., *Electrostatic self-assembly of macroscopic crystals using contact electrification,* **Nature Materials 2,** 241–245 (2003)

Abstract

1) Context: Self-assembly of components larger than molecules into ordered arrays is an efficient way of preparing microstructured materials with interesting mechanical and optical properties.

2) GAP: Although crystallization of identical particles or particles of different sizes or shapes can be readily achieved, the repertoire of methods to assemble binary lattices of particles of the same sizes but with different properties is very limited.

3) Purpose: This paper describes electrostatic self-assembly of two types of macroscopic components of identical dimensions using interactions that are generated by contact electrification.

4) Methodology: The systems we have examined comprise two kinds of objects (usually spheres) made of different polymeric materials that charge with opposite electrical polarities when agitated on flat, metallic surfaces.

5) Results: The interplay of repulsive interactions between like-charged objects and attractive interactions between unlike-charged ones results in the self-assembly of these objects into highly ordered, closed arrays. Remarkably, some of the assemblies that form are not electroneutral—that is, they possess a net charge.

6) Conclusions: We suggest that the stability of these unusual structures can be explained by accounting for the interactions between electric dipoles that the particles in the aggregates induce in their neighbors.

1) Context: Self-assembly of components larger than molecules into ordered arrays is an efficient way of preparing microstructured materials with interesting mechanical and optical properties.

2) GAP: Although crystallization of identical particles or particles of different sizes or shapes can be readily achieved, the repertoire of methods to assemble binary lattices of particles of the same sizes but with different properties is very limited.

3) Purpose: This paper describes electrostatic self-assembly of two types of macroscopic components of identical dimensions using interactions that are generated by contact electrification.

4) Methodology: The systems we have examined comprise two kinds of objects (usually spheres) made of different polymeric materials that charge with opposite electrical polarities when agitated on flat, metallic surfaces.

5) Results: The interplay of repulsive interactions between like-charged objects and attractive interactions between unlike-charged ones results in the self-assembly of these objects into highly ordered, closed arrays. Remarkably, some of the assemblies that form are not electroneutral—that is, they possess a net charge.

6) Conclusions: We suggest that the stability of these unusual structures can be explained by accounting for the interactions between electric dipoles that the particles in the aggregates induce in their neighbors.

Learning by doing

4th step:

practice the identification of distinct ways to convey rhetorical messages

Example: This ……. ……. the ……

	paper	addresses	problem
	letter	analyzes	case

Learning by doing

5th step:

The goal here is *to begin filling in* the content of your scientific paper with your own material.

in this process, try to enrich the possibilities by selecting other expressions (2nd step)

keep practicing identification of messages (4th step)

Learning by doing

6th step:

start all over again with the selected expressions, now classifying them according to rhetorical messages (e.g. describe, contrast, confirm, define, compare, introduce, etc).

the idea is to have a collection of expressions to be retrieved as you wish to analyze specific contents

keep selecting further expressions and identifying the messages

Learning by doing

7th step:

start working with full text passages, rather than only with separate sentences

repeat the procedures of Step 5.

now is the time to learn using connectives efficiently. Compile a list of expressions with *however, in contrast, indeed, on the other hand, furthermore, nevertheless, since, because*, etc.

Learning by doing

8th step:

it is time to produce a full section of a paper

select the subcomponents

implement them by using material from your earlier practices

check the use of connectives and the t ext coherence

9th step: Editing the text

Check the section for typos and other surface errors

Eliminate unnecessary words

Check the consistency of the subcomponents and their inter-relationship.

Analyze the contents for completeness and accuracy

Learning by doing

Important Note:

Though the AMADEUS strategy was not conceived as a tool to learn English, users may improve their proficiency by practicing with the language in context. For example, the material compiled should be excellent source for checking use of prepositions, phrasal verbs, connectives and even vocabulary of the field in focus.

Did you find *Learning by doing* tedious?

Too many steps? Too many expressions?

Too much reading? Too much homework?

But then, once you know them all, you will have become a very competent writer. We cannot offer any further help!

Disclaimers

1. Over the years I have seen people benefiting enormously from the AMADEUS strategy, but in all cases it was the practice and dedication of the user that made it a success.

2. This strategy cannot replace solid learning of English, and only works for users with reasonably good reception of English

Chu - March 2006