

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

Osvaldo N. Oliveira Jr.

**Núcleo Interinstitucional
de Linguística Computacional**

**USP-São Carlos, UFSCar,
Unesp-Araraquara**

<http://nilc.icmc.usp.br>

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 language; word by word translation; over-long/over-complex sentences.
- **Textual**
 - Use of rhetorical structures or strategies of the mother language; misuse of logical relations between sentences or phases; lack of references.

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

Major difficulties in writing

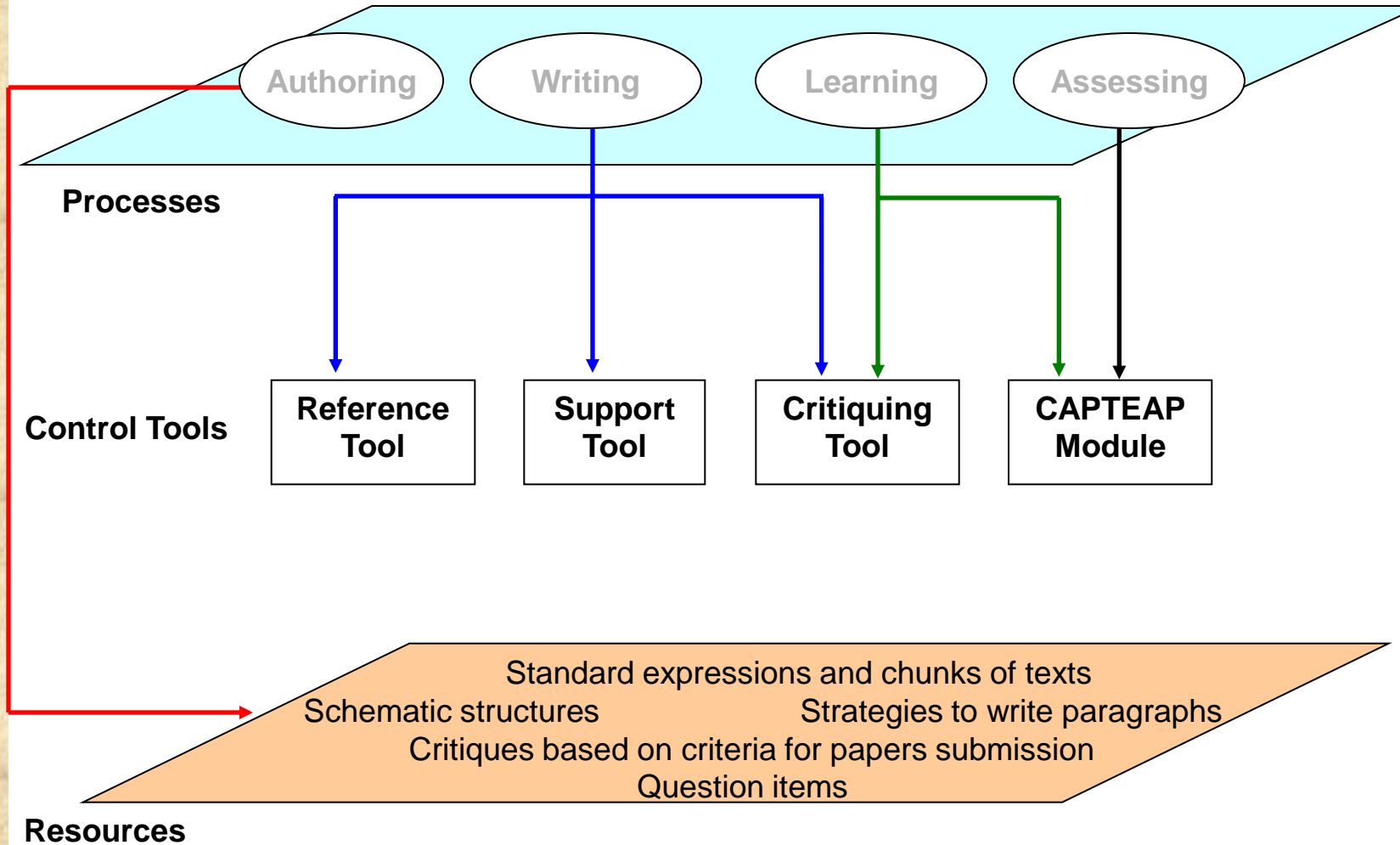
- **Surface errors (typos and grammatical errors)**
- **Long sentences and short, inadequate paragraphs**
- **Excessive number of unnecessary words**
- **Problems in text cohesion, inadequate use of markers and “zig-zag” in the discussion**
- **Lack of coherent “story” for the text**

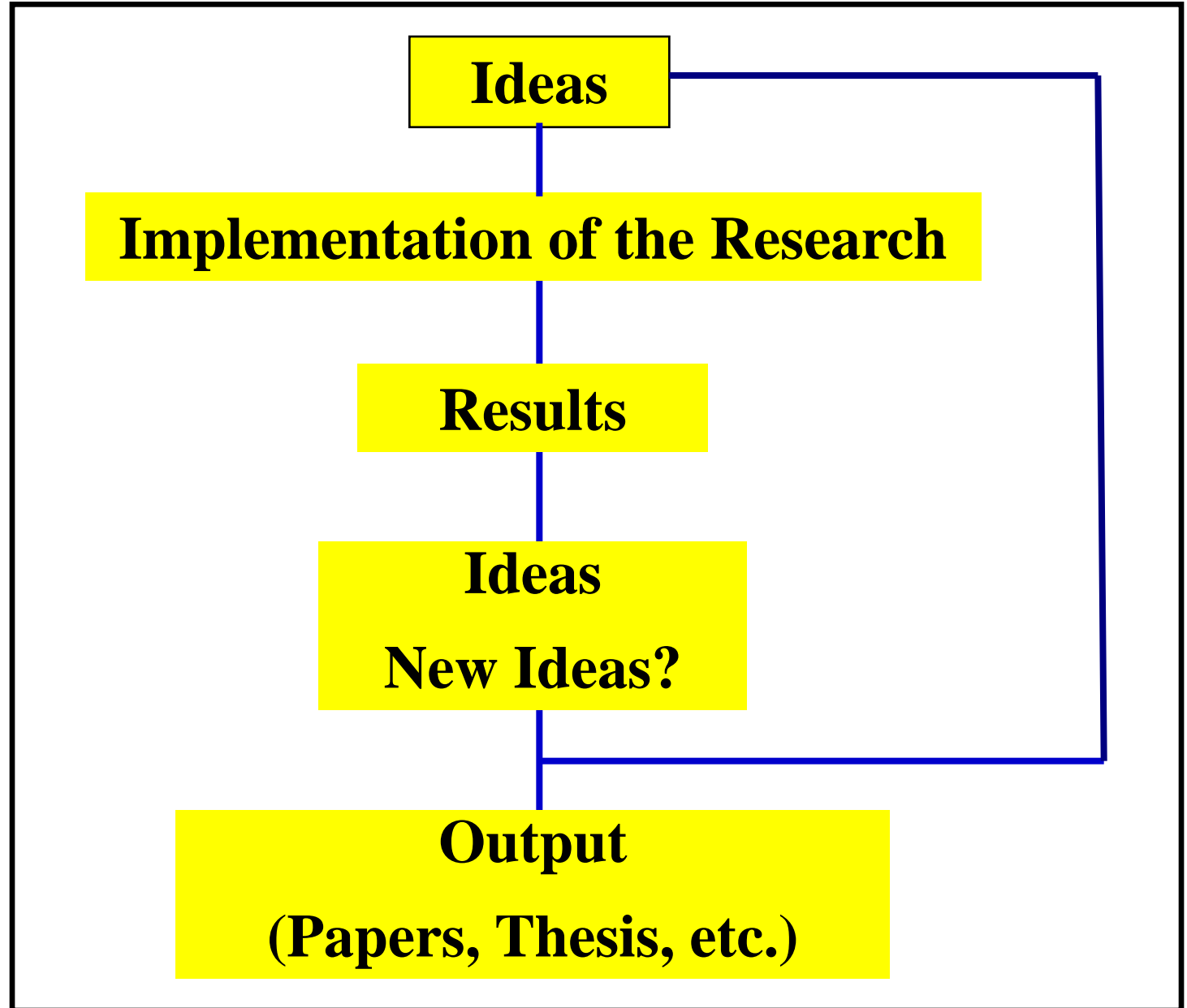
Corpus Linguistics - learning by example



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

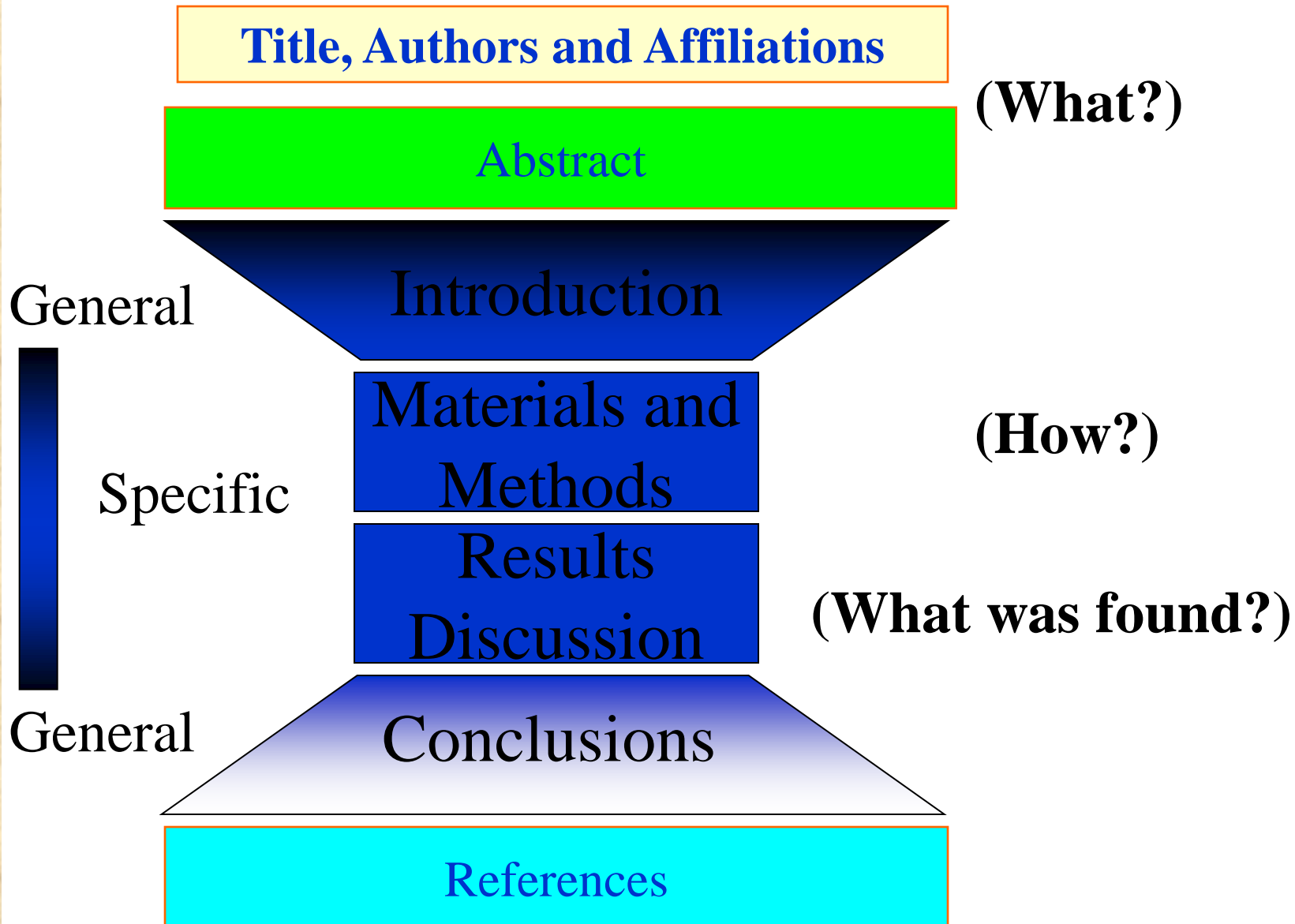
Resources & Tools in AMADEUS





**Scientists publish ideas and
concepts, NOT results!**

Overall Organization of a paper



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

The AMADEUS strategy

Aluísio & Oliveira, 1995

(www.nilc.icmc.usp.br)

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

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.

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..

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

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

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

4th step:

practice the identification of distinct ways to convey rhetorical messages

Example: This the
paper addresses problem
letter analyzes case

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)

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

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.

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 text 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

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