Journal of English for Academic Purposes 11 (2012) 357–366

Contents lists available at SciVerse ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Verb form indicates discourse segment type in biological research papers: Experimental evidence

Anita de Waard [a,*], Henk Pander Maat [b]

[a] Elsevier Labs, 71 Hanley Lane, Jericho, VT 05465, USA
[b] Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK Utrecht, The Netherlands

Keywords:
Discourse analysis
Scientific discourse
Genre analysis
Research articles

A B S T R A C T

Corpus studies suggest that verb tense is a differentiating feature between, on the one hand, text pertaining to experimental results (involving methods and results) and on the other hand, text pertaining to more abstract concepts (i.e. regarding background knowledge in a field, hypotheses, problems or claims).

In this paper, we describe a user experiment that investigates whether for biological readers, this tense correlation has a psychological correlate. To study this, we defined seven distinct discourse segments types and modified them either by changing the verb tense/mood (for all segment types), negation (for Problems), or presence of an epistemic matrix clause ('These results suggest…') for Implications.

Regardless of the original segment type, we found that for Facts, Results and Hypothesis segments, present tense yielded more Fact classifications, past tense more Result interpretations, and modal auxiliaries more Hypothesis interpretations. Methods statements were less sensitive to verb form. Problem segments required negations to be recognized, while Implications required introductory matrix clauses.

1. Introduction

Corpus studies (e.g. Burrough-Boenisch, 2003; Nwogu, 1997; Malcolm, 1987; Salager-Meyer, 1992; Schramm, 1996; de Waard & Pander Maat, 2009) suggest that verb tense is a differentiating feature between, on the one hand, text pertaining to experimental results (involving methods and results) and on the other hand, text pertaining to more abstract concepts (i.e. regarding known facts in a field, hypotheses, problems and implications). In particular, discourse segments pertaining to experimental methods and results are usually found to be in the past tense, whereas discourse pertaining to conceptual entities, such as statements of fact, implications or conclusions and hypotheses, are mostly given in the present tense. For instance, Li and Ge (2009) find that the majority of Nwogu's (1997) Moves 4, 5 and 6, which correspond to Methods, are in the past tense in the native-English corpus, whereas Moves 1 and 11, which correspond to facts and implications, are mostly stated in the present. Dahl (2009) finds that "new claims are often presented in the simple present in economics RAs". For hypotheses, the essential verbal form seems to be the use of modal auxiliary verbs such as '*could/may/might*'. Goal segments (which we do not explicitly investigate in the current study, for details, see below) are mostly given by means of *to*-infinitives: *'To investigate this issue…'*.

* Corresponding author.
E-mail addresses: a.dewaard@elsevier.com (A. de Waard), H.L.W.PanderMaat@uu.nl (H. Pander Maat).

1475-1585/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jeap.2012.06.002

These tense associations are supported by style guides for writing research articles. For example, Bem (2003) states:

> "Use the past […] tense when reporting how you conducted your study ("*Observers were posted behind…*") and specific past behaviors of your participants ("*Two of the men talked…*"). Use the present tense for results currently in front of the reader ("*As* Table 2 *shows, the negative film is more effective …*") and for conclusions that are more general than the specific results ("*Positive emotions, then, are more easily expressed when…*").,"

Similarly, the APA Manual (APA, 2001) says:

> "Use past tense (e.g., "*anxiety decreased significantly*") to describe the results. Use the present tense (e.g. "*the results of Experiment 2 indicate*") to discuss the results and to present the conclusions. By reporting conclusions in the present tense, you allow readers to join you in deliberating the matter at hand."

In this paper, we wish to investigate whether these correlations between verb tense (and mood) and discourse function are also psychologically based: in other words, when we change the tense of a clause, does its interpretation change as well? Specifically, we wish to investigate if experimental clauses, when presented in the present, read as facts; conversely, is a fact clause, presented in the past or with a modal auxiliary, interpreted as an experimental result or a hypothesis?

We believe it is essential to perform reader experiments to supplement corpus studies, which, though very valuable, cannot separate out the functions of tense and propositional content. When we want to claim that tense – or any other linguistic feature – 'marks' certain discourse segment types, we need to make clear that such a feature is actually a factor of how readers determine segment types. The most straightforward way to investigate 'marking'-relations is to change the markers while keeping the rest of the information constant, and check whether this manipulation affects the interpretation in the expected direction. That is what we have done here, by presenting clauses from full-text research articles in a different tense than the one in which they were written, and studying how they were interpreted by a group of domain specialists. Our overall hypothesis is that we should see a significant change in the frequency of interpretations associated with a different tense.

Before we describe this experiment, however, we offer a brief introduction to the discourse segment types by which we annotated our content.

2. Discourse segment types

The analysis that we use takes place at the level of the smallest unit of text with a single communicative intent: the discourse segment. Degand and Catherine (2005) define Minimal Discourse Units (MDU) as

> *"the smallest interactionally relevant complete linguistic unit[s], in a given context, that is constructed with syntactic and prosodic resources within their semantic, pragmatic, activity-type-specific, and sequential (…) context".*

A similar distinction is proposed by Polyani (1988) who defines discourse segments as:

> *"the syntactic constructions that encode a minimum unit of meaning of discourse function, interpretable relative to a set of contexts".*

There are many ways to chunk a text – see e.g. Marcu (2000) for an overview. Following a linguistic analysis in accordance with e.g. Marcu (2000) and Pander Maat (2002), we decided on the following criteria for segmenting text into non-overlapping segments with a single communicative intent. In Table A in the Appendix, we provide a summary of our segmentation decisions. In summary, we divide the text into clauses with a single (finite, or in some cases non-finite) verb in it; the only clauses we do not segment out are clauses that act as a Subject and restrictive relative clauses.

As an example, see sentences (1)–(3), taken from Voorhoeve et al. (2006):

(1) [An] escape from oncogene-induced senescence is a prerequisite for full transformation into tumor cells. (FACT)
(2) a. To identify miRNAs that can interfere with this process (GOAL)
b. and thus might contribute to the development of tumor cells, (HYPOTHESIS)
c. we transduced BJ/ET fibroblasts with miR-Lib (METHOD)
d. and subsequently transduced them with either RASV12 or a control vector (Figure 2B). (METHOD)
(3) After 2 or 3 weeks in culture, senescence-induced differences in abundance of all miR-Vecs were determined with the miR-Array. (RESULT)

Clearly, several distinct meanings can be stated within a single sentence in biology papers. For example, in sentence 2, the goal of the (sub)-experiment is first stated, followed by a hypothesis. After the comma, this is followed by a description of methods used. This is a significantly smaller granularity than e.g. the argumentative zones defined by Teufel, Carletta, and Moens (1999), or the sentences (Liakata, 2010) or bio-events (Nawaz, Thompson, & Ananiadou, 2010) used for computational linguistic work pertaining to scientific discourse analysis. However, the choice of granularity in computational work is determined by the tools at hand – as of yet, there seems to be no reliable way to automatically split a sentence into clauses. Table 2 states the level of granularity of our markup compared to four other systems.

Our segment type definition is given in brackets, following each segment. By analyzing a collection of biological research and classifying clauses with a similar communicative intent, we have arrived at a taxonomy of seven primary discourse