1 Purpose and aims

Multiword expressions (MWEs) are known to challenge various NLP applications such as machine translation and syntactic parsing. This paper deals with one example of an MWE, a construction that I will refer to as the Position Verb and Verb construction, which is found in all three Scandinavian languages, Norwegian Swedish and Danish, as well as in a number of other languages (see e.g., Kuteva 1999). The construction is made up of a position verb (PV) like ‘sit’, ‘lie’, ‘stand’, or ‘walk’, a conjunction ‘and’, and another verb. There is general agreement that the construction expresses some aspectual meaning apart from predicating the action described by the second verb (sometimes called the content verb). Examples of the construction in the three languages can be seen in Example 1, 2, and 3 below:

(1) *Jag står och fiskar.* (Swedish)
    ‘I am fishing’ (lit. ‘I stand and fish’)

(2) *Han sitter og skriver.* (Norwegian)
    ‘He is writing’ (lit. ‘He sits and writes’)

(3) *Jeg går og hoster.* (Danish)
    ‘I cough (a lot these days)’ (lit. ‘I walk and cough’)

The aim of the paper is to provide a linguistic description if the construction which can then possibly be used in the design of language technology applications to improve the linguistic quality of the end result.

2 Survey of the field

The PV and V construction has been described from different points of view as either pseudocoordination (Bjerre and Bjerre 2007a,b; Hilpert and Koops 2008; Lødrup 2002), auxiliation (Kuteva 1999), or more generally as a subtype of complex predication (Hilpert and Koops 2008) (this paper will focus on the last analysis and characterise the construction as a light verb construction which is
a type of complex predicate). In the following paragraphs, I will give a short overview of some of the central papers dealing with the $PV$ and $V$ construction in general and in the three Scandinavian languages specifically.

Kuteva (1999) describes the construction ‘sit’/‘stand’/‘lie’ + and + main verb as a way to express aspectual distinctions (continuative/durative/progressive) in Bulgarian, the Scandinavian languages, and a number of non-indoeuropean languages, and she describes how this use of the construction may have grammaticalised from an earlier, more literal use. Kuteva 1999 describes the construction as a case of auxiliation.

Bjerre and Bjerre (2007a,b) describe the syntactic properties of the version of the $PV$ and $V$ construction with sidde (‘to sit’) in Danish. They focus on the status of the construction as a hybrid phrase expressing both coordinating and subordination.

Hilpert and Koops (2008) present a diachronic perspective on the sitta (‘to sit’) version of the Swedish construction. In their analysis, the construction is an example of complex predication, more specifically a light verb construction (and so they disagree with Kuteva 1999). In their paper, they also argue for a monoclausal interpretation of the construction calling it a “monoclausal construction with two verbal heads” (Hilpert and Koops 2008, 244).

Lødrup (2002) describes the construction from a Lexical-Functional Grammar perspective. Contrary to Hilpert and Koops (2008) and Bjerre and Bjerre (2007a,b), Lødrup (2002) argues for a biclausal interpretation of the construction. He treats a wider range of constructions than that considered in this paper, but the $PV$ and $V$ construction he characterises as what he calls control constructions.

While the papers presented above differ in their focuses, theoretical backgrounds and to some extent in their conclusions, they seem to agree on a superficial characterisation of the construction in question, and I will briefly recount this characterisation here. On the form side, the construction is made up of one position verb and one other verb, and these must have the same value for tense and finiteness. The two verbs in the construction furthermore share the same overt subject. The construction can be modified for time or place but the situations described by the two verbs must take place at the same temporal and spatial location.

There is some disagreement as to the degree of semantic bleaching of the literal sense of the PV, and it is possible that the different PVs have retained their literal semantic meaning to different degrees. The different papers do seem to agree, however, that the meaning of the construction includes aspectual meaning of the type durative or progressive. The present project will aim to refine the description of both the form and semantics of the construction and make comparisons between the different PVs as well as between the three languages.

3 Programme description

In the paper, I will attempt to describe the $PV$ and $V$ construction in as theory neutral terms as possible (sometimes referred to as Basic Linguistic Theory) so that the description might more easily be adapted to fit different practical purposes within different theoretical frameworks.
To show how a linguistic description like this one can be relevant to the design of NLP applications, I will analyse the output of one such, namely Google Translate, when presented with examples of the \textit{PV and V} construction. Below is a preliminary table of contents for the paper as well as a tentative time plan for the development of the project.

\textbf{Outline of the paper:}

1 Introduction
2 The relevance of linguistic description to language technology
3 The \textit{PV and V} construction
   3.1 Form
   3.2 Semantics
4 Conclusion

\textbf{Time plan:}

Week 43: \textit{project proposal due}
Week 44-46: chapter on the construction
Week 47-49: chapter on relevance to language technology
Week 50: introduction and conclusion, \textit{first draft due on Friday}
Week 51: \textit{comments on classmates’ papers due on Friday}
Week 52-1: revision
Week 2: preparation for presentation, \textit{final deadline (Wednesday)}

\section{Significance}

Although there is a fair amount of research on how to deal with MWEs for the purposes of language technology, this specific light verb construction has (to my knowledge) not been investigated in this perspective. And while constructions of this kind are usually highly language specific, this construction exists in all three Scandinavian languages and has what appears to be closely related counterparts in a number of other languages (including Bulgarian, Afrikaans, and a number of other languages including non-indoeuropean languages, see Kuteva 1999; Kjeldahl N.d.; Lødrup 2002). This makes a description of this construction interesting as it builds a platform for descriptions of the related constructions in other languages to build on.

That the aspectual implications of the construction can pose difficulties for an NLP application like machine translation can be witnessed by even a superficial analysis of the output of a Google Translate translation of examples like Example 1 above. The Swedish sentence \textit{Jag står och fiskar} is translated literally into the English sentence ‘I stand and fish’, and while some alternative translations are given, the one that comes closest to conveying the aspectual semantics of the Swedish original is clumsy and grammatically incorrect: ‘I am standing and fish’. I argue that this calls for an in-depth description of the construction that can shed light on both is formal properties (for identification and extraction purposes) as well as it semantic properties (for interpretation and translation purposes).
References


