PAWS: Parser and Writer for Syntax

Drafting Syntactic Grammars in the Third Wave

Cheryl A. Black
H. Andrew Black

SIL International and University of North Dakota
2009

SIL Forum for Language Fieldwork 2009-002, May 2009
Copyright © 2008 Cheryl A. Black, H. Andrew Black and SIL International
All rights reserved

Contents

Abstract

Writing a grammar for a language can be a daunting task. Where does one begin? What are the most important things to include? While few of us will doubt the importance of having grammars for the languages and for the people we work with, it can seem like an inordinate amount of work.

Simons and Black (2009) note that we need to "work smarter" rather than work harder. One of the key notions is to produce "actionable" knowledge. They also note the role that XML technologies can play in enabling us to "work smarter." The **PAWS Starter Kit** takes advantage of these technologies to create "actionable" knowledge which can be turned into a draft of a grammar write-up. The user merely answers questions presented via web pages and, at the press of a button, the program produces a draft of a grammar write-up reflecting those answers.

1 Introduction

The task of writing a grammar for a language can seem so overwhelming that many give up without even trying. Clearly, having grammars for the languages and for the people we work with is important, but it is difficult to know where to start and what to include. Completing and revising a draft is a much less daunting task, and the **PAWS Starter Kit** can provide that needed head start.

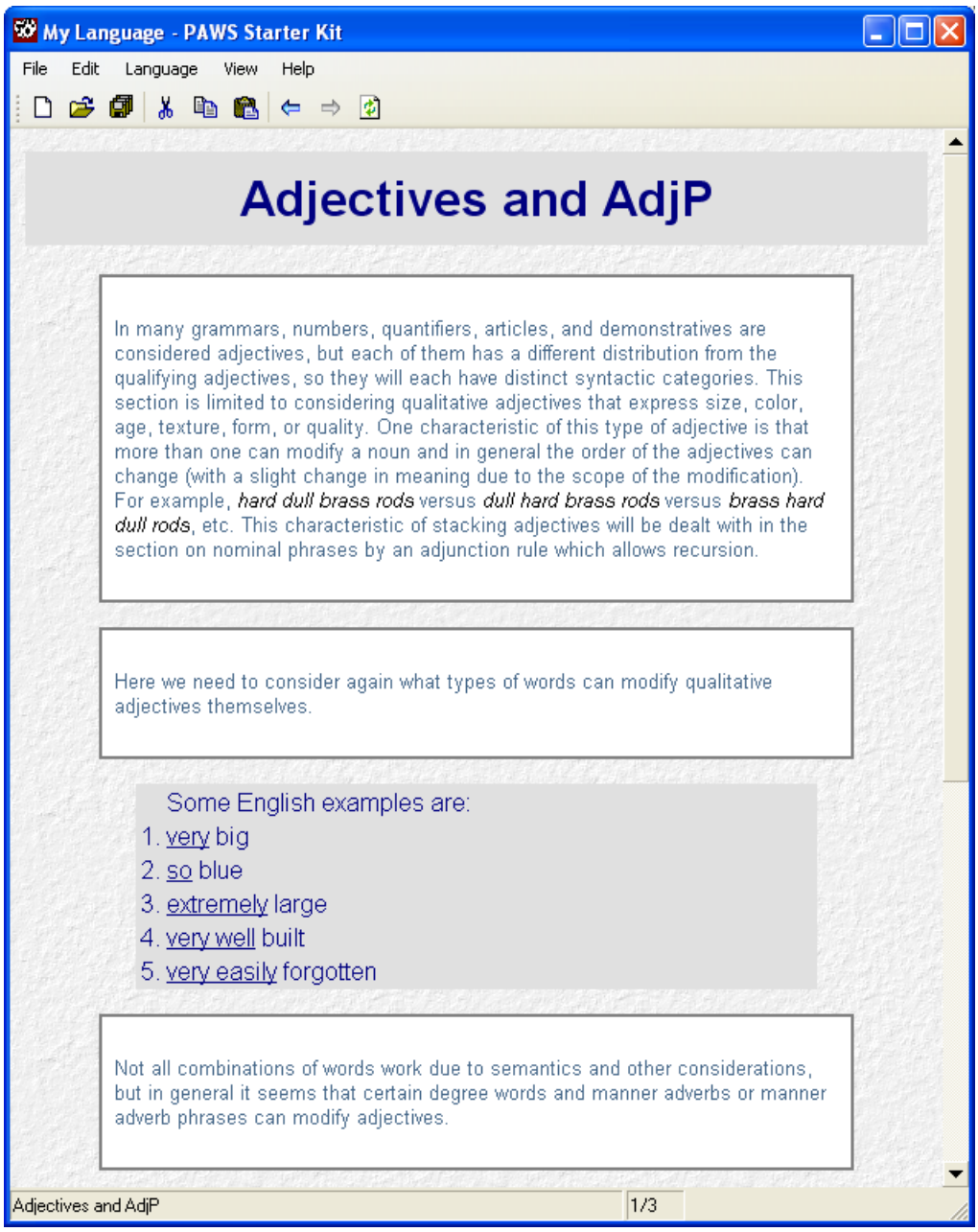
Simons and Black (2009) note that we need to "work smarter" rather than work harder. One of the key notions is to produce "actionable" knowledge. They also note the role that XML technologies can play in enabling us to "work smarter." The **PAWS Starter Kit** takes advantage of these technologies to create "actionable" knowledge which can be turned into a draft of a grammar write-up as well as a draft of a grammar file that can be used by the **PC-PATR** syntactic parsing program (see McConnel 1995). If you are not interested in doing syntactic parsing, you can merely ignore that part of the **PAWS Starter Kit** and just work on producing the written grammar draft.

The **PAWS Starter Kit** uses Unicode so it can handle most language scripts. It also works with Keyman 6.

The **PAWS Starter Kit** consists of a series of web pages that explain and illustrate the syntactic issues to be covered by the section and then has a series of multiple-choice questions about what happens in the language the user is studying. The user is also requested to provide example sentences and words.

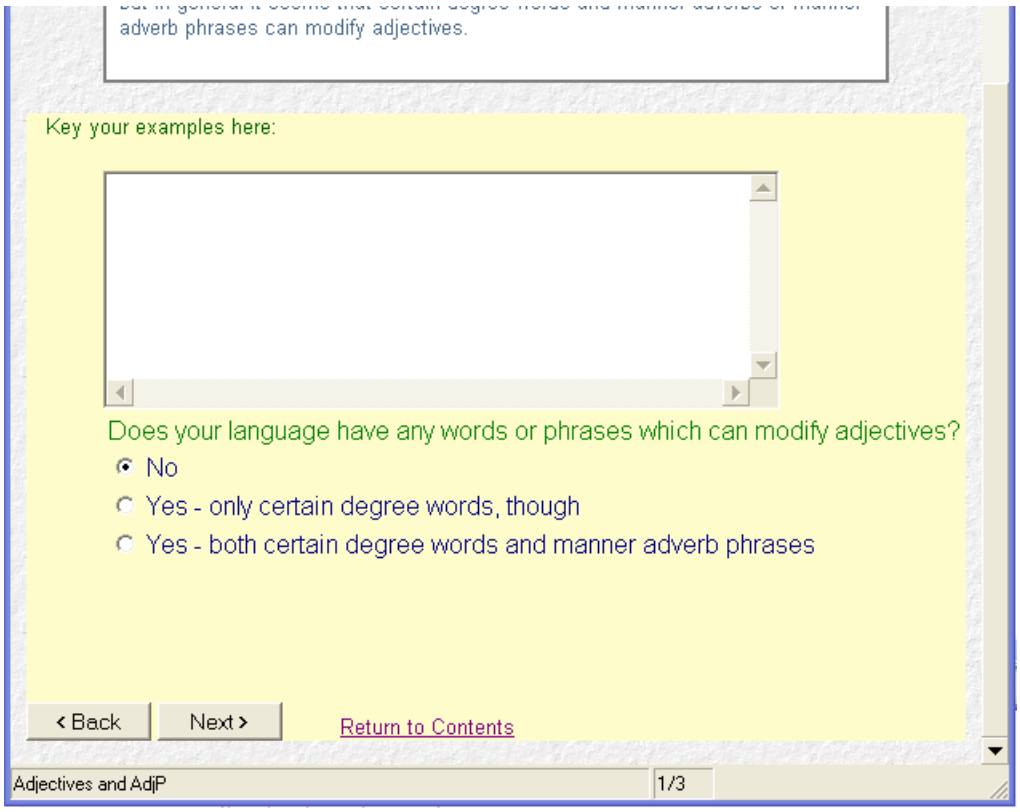
For example, consider the opening page on adjectives and adjective phrases. It begins as shown in (1).

(1)



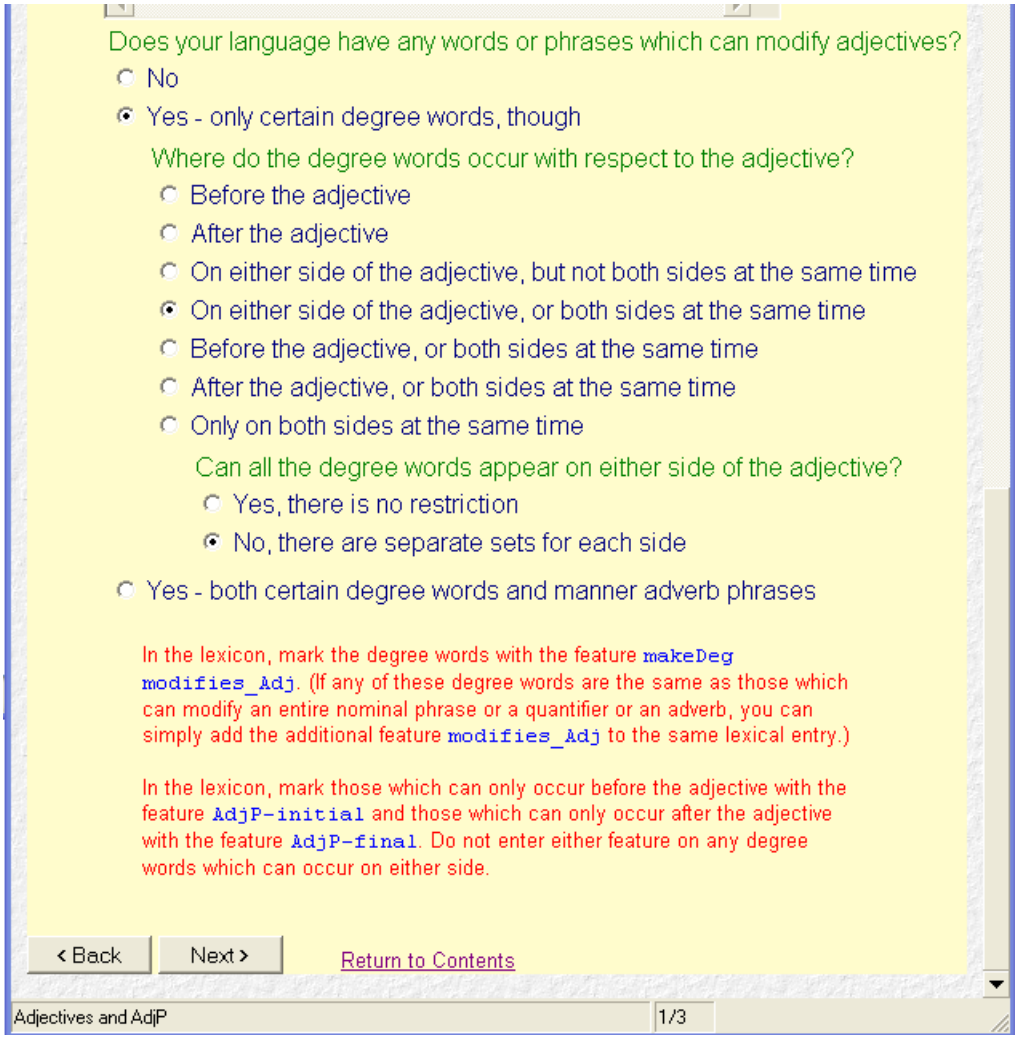
This explains and illustrates what this major section is about. Scrolling down, we see the rest of this page.

(2)



Notice that there is a place to key examples from the language of study and there are also multiple choice questions. The number of questions depends on how previous questions have been answered. For example, if we select the second choice ("Yes - only certain degree words, though"), we might get the following:

(3)



There are now additional questions to answer. In addition, there are instructions in red telling you how to enhance a **PC-PATR** lexicon[1] to be able to use the automatically generated **PC-PATR** grammar. If all you are interested in is the automatically generated grammar write-up, then you can ignore all the instructions that are in red.

When you are done with one page, you click on the Next button and go on to the next page.

By merely working one's way through the **PAWS Starter Kit**, one can relatively quickly obtain a draft of a grammar write-up. Thus, an approach to grammar writing like the **PAWS Starter Kit** can greatly diminish the size of the task.

In the rest of this paper, we briefly describe the scope of coverage in section 2, the nature of the outputs produced in section 3, how to customize the output in section 4, and some details of how the **PAWS Starter Kit** actually works in section 5. We end with describing how one can obtain the **PAWS Starter Kit** in section 6.

[1] Alternatively, one can use an **AMPLE** dictionary; see Weber, Black, and McConnel (1988), AMPLE, or CARLAStudio.

2 Contents

With respect to the scope of coverage, the **PAWS Starter Kit** consists of seventeen sections, some of which contain sub-sections. These seventeen main sections are:

1. Overall Typology Questions
2. Quantifiers and QP
3. Adverbs and AdvP
4. Adjectives and AdjP
5. Simple and Possessed Nominal Phrases
6. Proper Names
7. Pronouns
8. Pre/Post-positional Phrases
9. Basic, Single-clause Sentences
10. Complement Clauses
11. Questions
12. Relative Clauses
13. Adverbial Clauses
14. Negative Constructions
15. Coordination Constructions
16. Focus and Topic Constructions
17. Exclamations and Greetings

The answer you give in the overall typology section provides default answers for the rest of the sections. You can change these default answers, of course, when you need to.

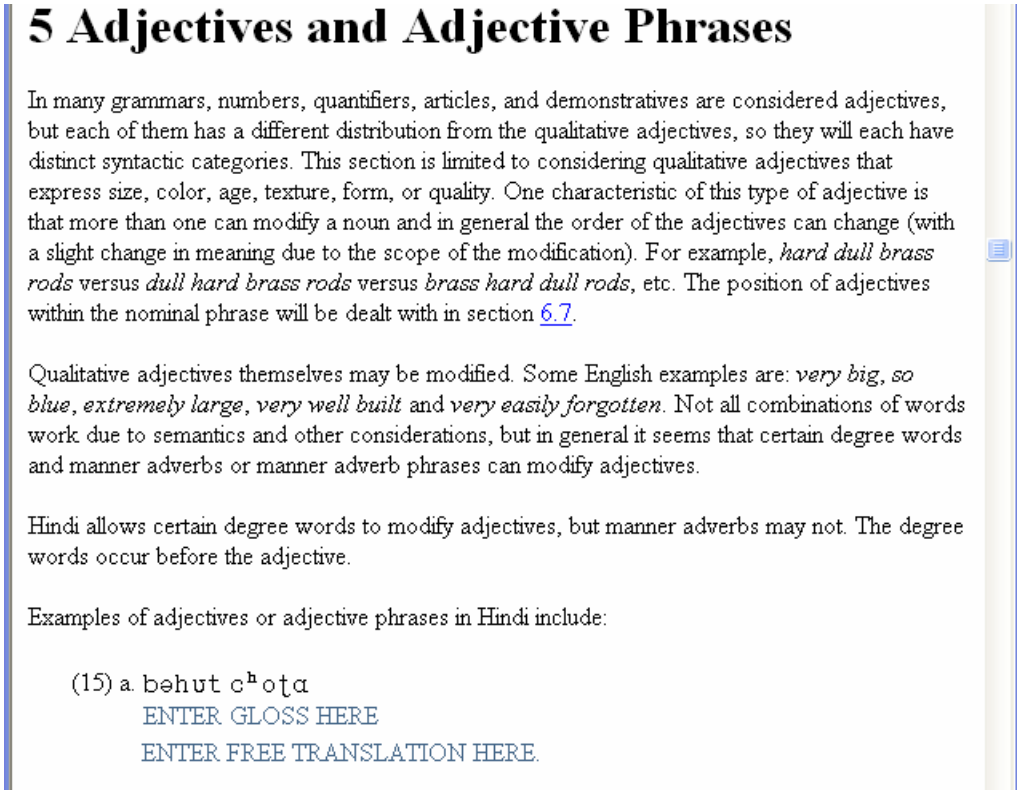
The order of the sections may not be what one might have expected if all that the **PAWS Starter Kit** did was to produce a draft of a grammar write-up. The reason it has this particular order is that one must deal with these issues in this order if one is building and testing a grammar for the **PC-PATR** syntactic parser.

3 Outputs

We now turn to what the **PAWS Starter Kit** actually produces. Whenever you move from one page to another, the **PAWS Starter Kit** automatically saves all of your answers to your "answer" file.

Every time you exit the **PAWS Starter Kit** or click on the save button, the **PAWS Starter Kit** generates a draft of the write-up. It also produces a draft of a **PC-PATR** grammar and sixteen test files to use in checking out the generated **PC-PATR** grammar. These files are too large to show here, but we will show in example (4) below a portion of one file that corresponds to the page shown in (1-3), where a user has entered a Hindi phrase using IPA in the examples section of (2). This example is from the writer output:

(4)



5 Adjectives and Adjective Phrases

In many grammars, numbers, quantifiers, articles, and demonstratives are considered adjectives, but each of them has a different distribution from the qualitative adjectives, so they will each have distinct syntactic categories. This section is limited to considering qualitative adjectives that express size, color, age, texture, form, or quality. One characteristic of this type of adjective is that more than one can modify a noun and in general the order of the adjectives can change (with a slight change in meaning due to the scope of the modification). For example, *hard dull brass rods* versus *dull hard brass rods* versus *brass hard dull rods*, etc. The position of adjectives within the nominal phrase will be dealt with in section 6.7.

Qualitative adjectives themselves may be modified. Some English examples are: *very big*, *so blue*, *extremely large*, *very well built* and *very easily forgotten*. Not all combinations of words work due to semantics and other considerations, but in general it seems that certain degree words and manner adverbs or manner adverb phrases can modify adjectives.

Hindi allows certain degree words to modify adjectives, but manner adverbs may not. The degree words occur before the adjective.

Examples of adjectives or adjective phrases in Hindi include:

(15) a. bəhut cʰoʈa
ENTER GLOSS HERE
ENTER FREE TRANSLATION HERE.

Notice that each example phrase entered will show up as an interlinear example, minus the appropriate data for the gloss and free translation. The writer output is in **XLingPaper** format. This can be easily edited with the **XMLmind XML Editor** (see Black 2009 or go to http://www.sil.org/~blacka/xlingpap/index.htm). Thus, one can add the missing information for glosses and free translations using this tool.

4 Customization

Besides editing the final results of the **PAWS Starter Kit**, it is also possible to customize the files of the starter kit itself to be used to produce a grammar with the chapters reordered or a writeup in a different language, for example. .

With respect to the writer output, the order of the sections and/or the exact mode of presentation may not be what you would prefer. There is a way that you can customize this that also uses a special set of configuration files for the **XMLmind XML Editor**. Using these, you can begin with what the **PAWS Starter Kit** comes with and then modify it to your liking. You generate a special file (called a transform) by pressing a button and then replace the original version of this file with the new one. The next time you run the **PAWS Starter Kit** and save the files, it will use your modified version of the writer transform. Contact Andy if you wish to try this.

Customizing the automatically generated **PC-PATR** grammar file is not so simple. Contact Cheri if you wish to try this.

5 Under the Hood

For those of you who are interested in such things, this section explains how all this works.

There are four sets of information, all of which use XML technologies (including a Document Type Definition for each one):

1. The structure of the answers to the questions (including the example sentences and words).
2. The content of each "question" web page, including conditions for when certain questions should or should not appear based on the answers the user has given so far.
3. The meta-authoring descriptions for the write-up, including conditions for when certain text should or should not appear based on the answers.
4. The meta-authoring descriptions for the PC-PATR grammar file, including when certain rules should or should not appear based on the answers.

A non-programmer "expert" in syntax (Cheri in our case) creates and maintains all four sets of information as XML files. These are then packaged together in a computer program that displays the "question" web pages, records the answers, and produces the draft documents.[2]

6 How to get the `PAWS Starter Kit`

The latest version of the **`PAWS Starter Kit`** can be obtained from http://www.sil.org/computing/catalog/show_software.asp?id=85.

References

Black, H. Andrew. 2009. "Writing Linguistic Papers in the Third Wave." *SIL Forum for Language Fieldwork* 2009-004. http://www.sil.org/silepubs/abstract.asp?id=52286

McConnel, Stephen. 1995. "PC-PATR Reference Manual." SIL International. (http://www.sil.org/pcpatr/manual/pcpatr.html).

Simons, Gary F. and H. Andrew Black. 2009. "Third Wave Writing and Publishing." *SIL Forum for Language Fieldwork* 2009-005. http://www.sil.org/silepubs/abstract.asp?id=52287

Weber, David J., H. Andrew Black, and Stephen R. McConnel. 1988. *AMPLE: A Tool for Exploring Morphology.* Occasional Publications in Academic Computing No. 12. Dallas, Texas: Summer Institute of Linguistics. http://www.sil.org/computing/catalog/show_software.asp?id=1

[2]Version 0.6 of the **`PAWS Starter Kit`** actually uses a generic shell tool that potentially could be used to package up similar applications, such as for phonology, say.