RESOLVING LR TYPE CONFLICTS
AT TRANSLATION OR COMPILE TIME

ABSTRACT.
The paper considers circumstances in which it is advantageous to resolve reduce-reduce conflicts at compile time, rather than at compiler-construction time. The application considered is that of translating English to one of the Romance languages, such as Italian where adjectives and nouns have distinctive forms depending on their gender.

As an example of the application area of the algorithm that this paper describes, consider the problem of natural language translation with respect to a restricted set of sentences, such as may occur in e.g. translation systems for tourists, voice-recognition systems, air-traffic control, dispatchers for trucks and taxis, police radio transmission, sport coaching, text messaging, lab results, and various medical and military contexts. Normal LR parser construction using attribute grammars are some of the tools employed for translation systems, but there are quite simple cases where constructing the complete parser in advance is not feasible. For the purposes of illustration, let us initially consider the problem of translating from English to Italian a trivially small set of sentences of the form:

The <adjective> student is a <adjective> <person>.

e.g. The Italian student is a tall man (Lo studente italiano é un uomo alto)
or The Italian student is a tall woman (La studentessa italiano é una donna alta)

All the translation engines available on the web that we have tried, including those provided by Google, Yahoo, etc, translate “The Italian student is a tall woman” partially or wholly in the masculine form (but will usually correctly translate “The tall woman is an Italian student”).

Restricting, for the sake of exposition, the set associated with the <adjective> qualifying “student” to {Italian, Austrian}, and set associated with the <adjective> qualifying “person” to {tall, good}, and that associated with <person> to {man, woman}, a possible grammar\(^1\) for such sentences, in which gender is specified using the prefixes “m_” and “f_” for the purpose of facilitating translation into Italian, is:

\[
\text{sentence} \rightarrow \text{m_the m_adjective-student IS m_a m_adjective-person} \\
| \text{f_the f_adjective-student IS f_a f_adjective-person}
\]

\[
\text{m_the} \rightarrow \text{the} \\
\text{f_the} \rightarrow \text{the}
\]

\[
\text{m_the-adjective-student} \rightarrow \text{m_adjective_1 m_student} \\
\text{f_the-adjective-student} \rightarrow \text{f_adjective_1 f_student}
\]

\[
\text{m_adjective_1} \rightarrow \text{Italian} \\
| \text{Austrian}
\]

\[
\text{f_adjective_1} \rightarrow \text{Italian} \\
| \text{Austrian}
\]

\[
\text{print(“Lo”)}
\]

\[
\text{print(“La”)}
\]

\[
\text{print(SavedAdjective)}
\]

\[
\text{SavedAdjective = “italiano”}
\]

\[
\text{SavedAdjective = “austriaco”}
\]

\[
\text{SavedAdjective = “italiana”}
\]

\[
\text{SavedAdjective = “austriaca”}
\]

\(^1\) with some code-generation (in pseudo-C) shown in the manner employed by YACC. SavedAdjective is a string variable.
The parser for this grammar is supplied in the following parsing machine:

The above machine contains a reduce-reduce conflict at state 1.
The grammar is a LR(6) one, and the contexts for \texttt{m\_the\rightarrow the} are:
- Italian student is a tall man
- Austrian student is a tall man
while the contexts for \texttt{f\_the\rightarrow the} are similar except for the replacement of “man” by “woman”.

\begin{verbatim}
m\_student \rightarrow student f\_student \rightarrow student
IS \rightarrow is
m\_a \rightarrow a f\_a \rightarrow a
m\_adjective-person \rightarrow m\_adjective\_1 man
f\_adjective-person \rightarrow f\_adjective\_1 woman
m\_adjective\_2 \rightarrow tall | good
f\_adjective\_2 \rightarrow tall | good

The grammar is a LR(6) one, and the contexts for \texttt{m\_the\rightarrow the} are:
- Italian student is a tall man
- Austrian student is a tall man
while the contexts for \texttt{f\_the\rightarrow the} are similar except for the replacement of “man” by “woman”.
\end{verbatim}
Note 1. Simple extensions to the above grammar provide sentences such as
the student is a man who is tall
in which “man” in this case is not the last word.
Note 2. For the purposes of the translation program, the English sentence provided is
assumed to be a valid member of some set S of sentences. The translation program is not
required to check that this is so. Hence the grammar employed may be one for a superset
of S, and include sentences which are not in S and are not expected to ever occur in
practice, such as:
the tall student is a short man
Note 3. The set S of sentences may include ones such as
the tall student is a happy person
in which the gender cannot be determined (and so the gender the translation makes use of
in these circumstances is arbitrary).
Note 4. The gender involved can be determined in some cases by the adjective employed.
For instance the gender involved may be taken to be male in:
the student is a brawny person
or:
the student is macho
while the female gender is implied by sentences such as:
the student is a buxom person
or:
the student is pretty

There are several thousand English adjectives. If we alter the above grammar to provide
for a selected thousand of them, then since adjectives occur at two places within the
the set of sentences, there will be a million contexts of length 6 for both m_the→the
and f_the→the in the resulting parsing machine. In addition there are thousands of
occupations besides “student”. It should be clear from the above illustrative example, fairly
simple subsets of English may be considered in which the number of contexts needed to
resolve reduce-reduce conflicts is immense, even immense enough to exceed the storage
capacity of present-day computers. Such a large number of contexts will also be obtained
where the number of choices at each point in the set of sentences is small, but the number
of places where there are choices is large. In the above cases, the number of contexts
rises exponentially with the product of the sizes of the sets of choices involved, but the
number of additional states that are generated rises only with the sum of these sizes.

Furthermore if we allow strings of adjectives, such as:
the student is a tall, thin, debonair man
and do not place an upper bound on the size of such strings, then the number of possible
contexts required to resolve reduce-reduce conflicts in the resulting parsing machine
becomes unbounded (and associated grammars may not be LR(k) for any k).

Without evaluating the contexts at state 1 at compiler-construction time (impractical in the
case where a much greater number of adjectives and occupations is provided for than that
in the above grammar), we can nevertheless employ the above machine to resolve the
reduce-reduce conflict in the cases described above by considering the actual context
obtained at translation time. Consider, for example, the sentence
The Austrian student is a tall woman
In the case where “m_the→the” is chosen at state 1, the parse of the sentence, using the
parsing machine would lead to state 2, and then (with next input symbol “Italian”) to state 3,
where the production “m_adjective1→Italian” would be carried out. This would lead to state 5, and then (with next input symbol “student”) to state 6 and production “m_student→student”. This would lead to state 7 and the production “m_adjective-student→m_adjective1 student”, leading to state 8, and then (with next input symbol “is” to state 9 and production “IS→is”, leading in turn to state 27, and then (with next input symbol “a”) to state 11. The production “m_a→a” follow, leading to state 12 and then (with next input symbol “tall”) to state 13. Here the production “m_adjective2→tall” would be carried out, leading to state 15. At this stage the next input symbol would be “woman”. But no action for “woman” is defined at state 15, showing that “Italian student is a tall woman” is not a valid context of length 6 of “m_the→the” at state 1. On the other hand, by a similar procedure to that described above, we can show that, if “f_the→the” is chosen at state 1, then the context “Italian student is a tall woman” will lead to state 33, showing that that context is a valid context of length 6 for “f_the→the” at state 1. This resolves the conflict at state 1. The above method of resolving this conflict, will remain unchanged, even if the grammar is augmented to (say) allow a thousand adjectives instead of only “Italian” and “Austrian”, and a thousand adjectives instead of only “tall” and “good”.

This is an example of the cases described above where the number of contexts required to resolve a reduce-reduce conflict at translation-construction time is an exponential function of the sizes of the set of choices involved, but the number of parsing machine states required is a function only of the sum of these sizes, and further the number of steps required to resolve the conflict at translation time is a function only of the number of such sets of choices.