



Universidade do Minho  
Escola de Engenharia  
Departamento de Informática

# Assessing Attribute Grammars' Quality: metrics and a tool

João Cruz<sup>1</sup>, Pedro Rangel Henriques<sup>1</sup> and Daniela da Cruz<sup>3</sup>

**SLATE'2015**

<sup>1</sup>Dpt.Informatica/Centro Algoritmi, Universidade do Minho, Braga, Portugal

<sup>2</sup>Centro Algoritmi, IPCA, Barcelos, Portugal

2015-June-19

# Context

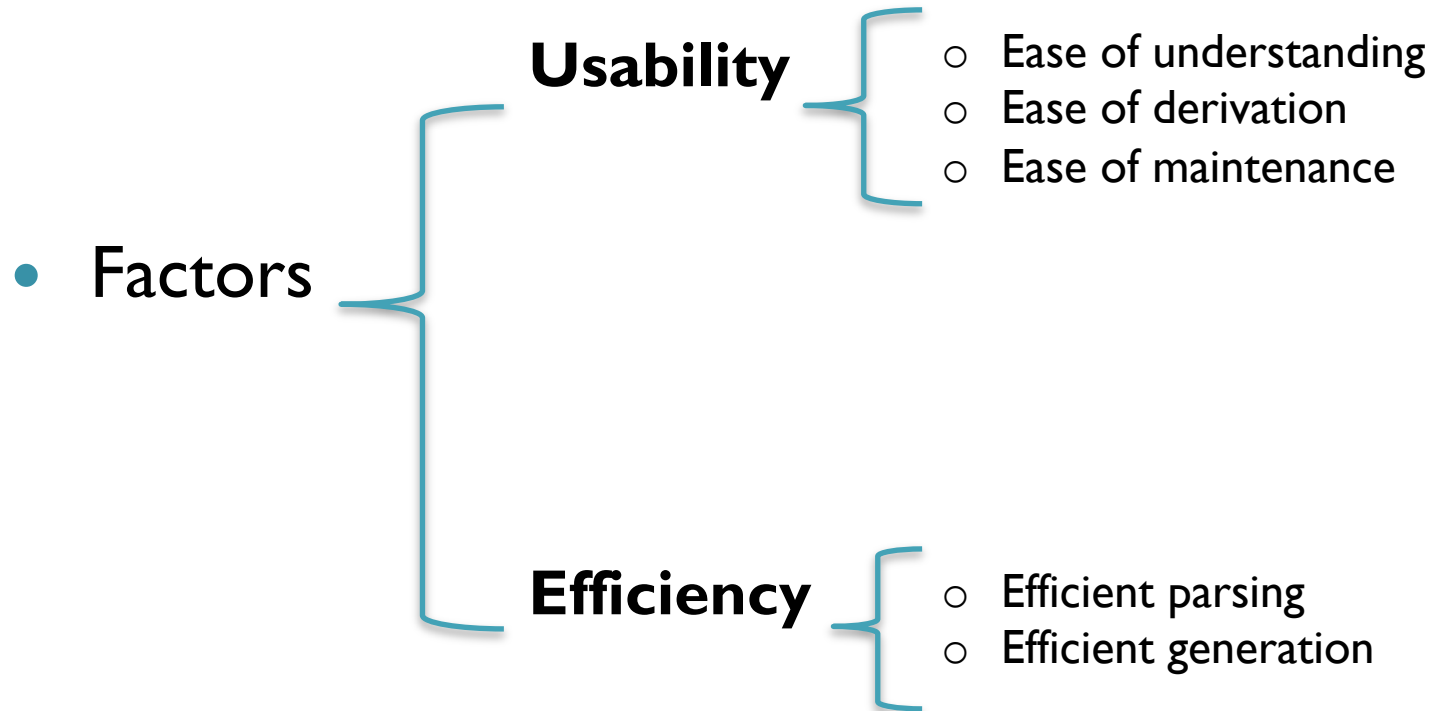
- The term engineering is used to apply mathematical, technical and scientific knowledge for the creation, maintenance and improvement of products
- The right use of this discipline leads to better results in terms of product quality and productivity
- To discuss product quality, engineers define metrics and measuring processes

# Motivation

- To reason about Grammars Quality aiming at
  - Improving it as a
    - specification document,
    - instrument in SW development.
  - Influencing the Language Quality
- To define metrics for CFG and AG
- To implement a tool
  - to automatize the measuring process
  - to help in assessing its quality and maintenance

# Grammar Quality

Considering a Grammar as a two fold formalism used to *define (generate) a language and guide the recognition of that language*



# Grammar Quality

**Usability** measures the level of satisfaction when using the grammar to *learn the language*, to *write sentences*, and to *debug or evolve* it.

- it has to do with symbol/attribute identifiers, the grammar size, and writing style

**Efficiency** measures the generated processor quality (the *parsing time* and the *size/complexity of the parsing tables*), and the performance of the generation process (*generation time* and of the *size of intermediate data structures*).

- is affected by external issues (methods, techniques and algorithms used), but it also depends on the size of the grammar and the writing style.

# Grammar Quality

It is important to identify **grammar characteristics** (able to be measured) that have a direct influence on the grammar quality:

- the Identifiers of Symbols or Attributes;
- the number of Symbols or Attributes, of Productions and Unit Productions;
- the length of the RHS (Right-Hand Side);
- the Notation and the Recursion schema used to write the Productions;
- the Attribute Types and simplicity of the Semantic Operators;
- the number of Semantic Rules (attribute evaluation rules, contextual conditions, and translation rules);
- the Attributive schema (purely synthesized, or mixed (inherited and synthesized));
- the Syntactic/Semantic Complexity (Symbol/Attribute Dependencies);
- the Modularity.

# Grammar Metrics

- Context Free Grammar Metrics

- ❖ Size Metrics

- ❖ Grammar Size: #T, #N, #P, #UP, #R, §RHS, §RHS-Max, §Alt, §Alt-Max, #Mod
    - ❖ Grammar Syntax Complexity: FanIn, FanOut
    - ❖ Parser Size: #RD, §TabLL, §DA-LR, §TabsLR

- ❖ Style Metrics

- ❖ Form of Recursion: Direct, Indirect or Mixed
    - ❖ Type of Recursion: Right, Left, LL(1) or Mixed
    - ❖ Notation: BNF or extended-BNF

- ❖ Lexicographic Metrics

- ❖ Clear identifiers for terminals and non-terminals
    - ❖ Clear reserved-words and signs
    - ❖ Flexibility of terminal-classes
    - ❖ Comment types: block, inline and meta-information inside block

# Grammar Metrics

- Attribute Grammar Metrics

- ❖ Size Metrics

- ❖ Grammar Size: **#A, #IA, #SA, #CR, #CC, #TR**
    - ❖ Grammar Semantic Complexity: **FanIn, FanOut**

- ❖ Style Metrics

- ❖ Attribute Complexity
    - ❖ Complexity of the Attributive Operations
    - ❖ Evaluation Scheme: form of aggregation pattern and form of attributive accumulation pattern
    - ❖ Semantic Restriction Scheme: **synthesized, right point, inherited or not typical pattern**
    - ❖ Translation Scheme: **synthesized, right point, inherited or not typical pattern**
    - ❖ Language Style and Specificity

- ❖ Lexicographic Metrics

- ❖ **Clear identifiers** for **attributes**
    - ❖ **Clear identifiers** for **attributive operators**



# Grammar Tool: GQE

- Helps on the assessment of grammar quality by performing the automatic evaluation of a large set of metrics
- Based on the metrics computed, any Grammar Engineer will easily be able to reason about the quality of its grammar and to improve it
- Will produce a Report on the grammar quality using a CBR system to accumulate past experiences

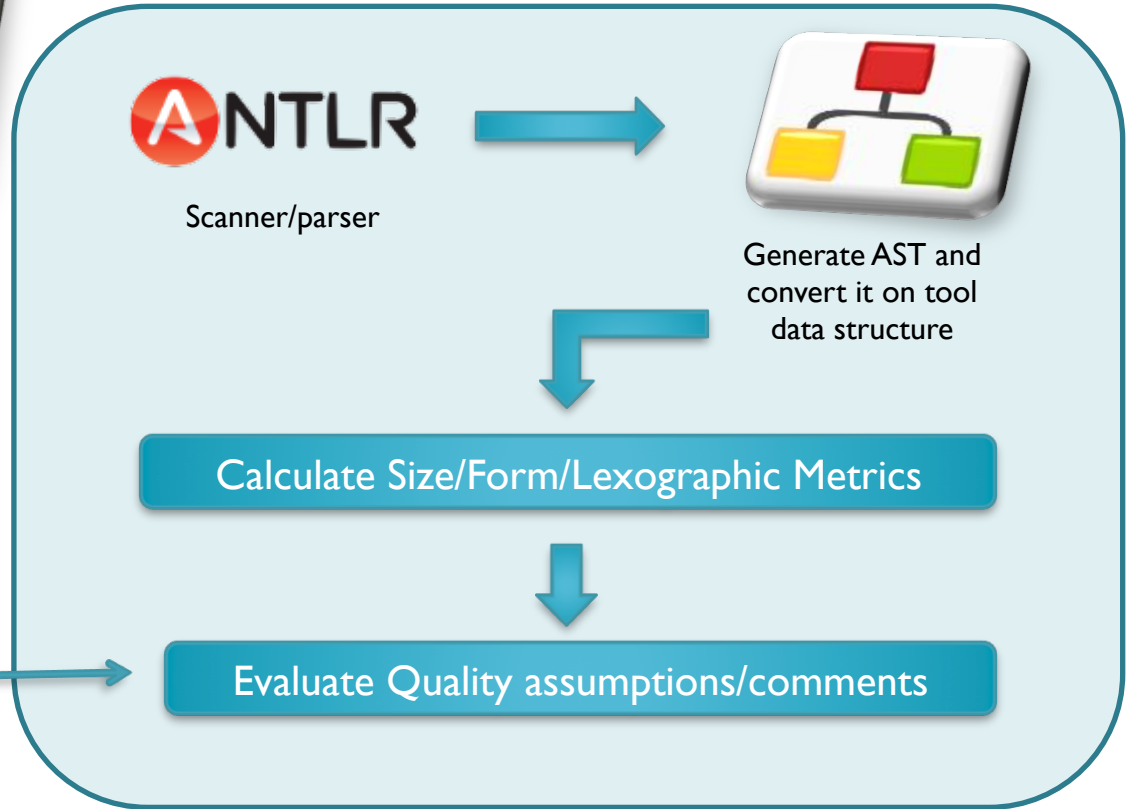
# GQE: Grammar Quality Evaluator

```
grammar binarip;  
  
bins  
: b1= bit{0}  
  (b2= bit[$b1.numero]{ $b1.numero = $b2.numero; })*  
  {System.out.println("Total: "+$b1.numero);}  
  ;  
  
bit [int anterior] returns [int numero]  
: NUM {$numero = ($anterior * 2) + $NUM.int;}  
  ;  
  
NUM : '0' | '1';  
  
Sep: ('\r'?'\\n'|' '\t')+ ->skip;
```

Input  
Context Free/  
Attribute Grammar



Repository of quality  
grammar/language  
characteristics



Output  
Quality Report

# GQE - implementation

Most of the metrics are evaluated by direct measuring (counting elements),  
but some of them are pretty difficult to evaluate

complex algorithms are needed:

- Dependency Graphs builder and analyzer
- Java code analyzer (to identify types, operations, ...)
- Identifier Splitter and Expander

# GQE - results

- Context-Free Grammar

## Lisp Grammar Example

```
grammar LispGIC;  
  
lisp  
    : sExp  
    ;  
  
sExpList  
    : sExp sExpList  
    |  
    ;  
  
sExp  
    : NUM  
    | WRD  
    | '(' sExpList ')'  
    ;
```

Size Metrics	
#T	4
#N	3
#P	6
#UP	0
#R	2
§RHS	1,3
§RHS-Max	3
§Alt	2,0
§Alt-Max	3
#Mod	0
FanIn	2,7
FanOut	1,1
§RD	7
§TabLL	15
§AD-LR	10
§TabsLR	50;30

# GQE - results

- Context-Free Grammar

## Lisp Grammar Example

```
grammar LispGIC;
```

```
lisp
```

```
    : sExp  
    ;
```

```
sExpList
```

```
    : sExp sExpList  
    |  
    ;
```

```
sExp
```

```
    : NUM  
    | WRD  
    | '(' sExpList ')  
    ;
```

### Style Metrics

Recursion Form	FormMixedRec
Recursion Type	Right Recursion
Notation	Pure-BNF

### Lexicographic Metrics

Clear Identifiers	5 / 5
Clear Keywords/ Signs	?
Flexibility Terminal-Classes	?
Comment Type	0

# GQE - results

- Context-Free Grammar  
Lisp Grammar Example

Identifier	Splitter/Expansion
lisp	lisp
sExpList	symbolic / expression / list
sExp	symbolic / expression
NUM	number
WRD	word

Style Metrics	
Recursion Form	FormMixedRec
Recursion Type	Right Recursion
Notation	Pure-BNF

Lexicographic Metrics	
Clear Identifiers	5 / 5
Clear Keywords/ Signs	?
Flexibility Terminal-Classes	?
Comment Type	0



# GQE - results

- Attribute Grammar

## Lisp Grammar Example

\* Not applicable to Antlr format grammars

Size Metrics	
#A	17
#AI	8
#AS	9
#CR	23
#CC	*
#CR	*
FanIn	0,81
FanOut	1,14

```
sExp[ int countN_in, int countW_in, int level_in, ArrayList<Pair> list_in]
  returns [int countN_out, int countW_out, ArrayList<Pair> list_out]
    : NUM    { $countN_out = $countN_in +1;
              $countW_out = $countW_in;
              $list_out = insere($list_in, $NUM.text, $level_in);
            }
    | WRD    { $countN_out = $countN_in;
              $countW_out = $countW_in +1;
              $list_out = insere($list_in, $WRD.text, $level_in);
            }
    | '(' sExpList[countN_in, countW_in, level_in+1, list_in] ')
      { $countN_out = $sExpList.countN_out;
        $countW_out = $sExpList.countW_out;
        $list_out = $sExpList.list_out;
      }
    ;
```

# GQE - results

- Attribute Grammar

## Lisp Grammar Example

\* Not applicable to Antlr format grammars

Lexicographic Metrics	
Clear Attributes Identifiers	17 / 17
Clear Attributive operators Identifiers	?

Style Metrics	
#Att. Complexity	5 / 17
#Att. Op. Complexity	?
Evaluation Scheme	?
Semantic Restriction Scheme	*
Translation Scheme	*
Language Style	Imperative
Language Specificity	Standard OO



# Conclusion

- project summary
  - A tool for Grammars Quality Assessment was introduced
  - The Metrics evaluated were listed
  - The rationale behind the definition of that set of Metrics was discussed based on Grammar Quality Factors
- present project status
  - Almost all the metrics are now evaluated
- lessons learned
  - Most of the metrics proposed can be evaluated automatically in a systematic and objective way
  - Some metrics must be evaluated by the user
  - It is not realistic to find a formula and compute a Quality Value

# Conclusion

- future work
  - Finish/improve the Metric Evaluation
  - Work out on the Interactive Interface
  - Developed the Case-based Reasoning system
  - Developed the Learning system to create a Quality Report
  - Study Language Quality and how grammar quality impacts on it