# JU-PTBSGRE: GRE Using Prefix Tree Based Structure

**Sibabrata Paladhi**
Department of Computer Sc. & Engg.
Jadavpur University, India
sibabrata_paladhi@yahoo.com

**Sivaji Bandyopadhyay**
Department of Computer Sc. & Engg.
Jadavpur University, India
sivaji_cse_ju@yahoo.com

## Abstract

This paper presents a Prefix Tree based model of Generation of Referring Expression (RE). Our algorithm PTBSGRE works in two phases. First, an encoded prefix tree is constructed describing the domain structure. Subsequently, RE is generated using that structure. We evaluated our system using Dice, MASI, Accuracy, Minimality and Uniqueness scoring method using standard TEVAl tool and the result is encouraging.

## 1 Introduction

Generation of referring expression (GRE) is an important task in the field of Natural Language Generation (NLG) systems. The existing algorithms in GRE lie in two extremities. Incremental Algorithm is simple and speedy but less expressive in nature whereas others are complex and exhaustive but more expressive in nature. We propose a new Prefix Tree (Trie) based framework for modeling GRE problems. It incorporates intricate features of GRE (like set and boolean descriptions, context sensitivity, relational description etc.) while achieving attractive properties of Incremental algorithm (simplicity, speed etc.). The prefix tree based algorithm works in two phases. First, it encodes the description, stored in the knowledge base, in the form of prefix tree structure. Secondly, it generates the referring expression identifying the target object, which is basically a node search problem in the tree. The edges in our encoded trie structure are labeled and the path from root to that node forms the distinguishing description for the target object.

Let **D** be the Domain, **r** be the target object and **P** be the '**PreferredAttributes'** List.The Trie constructionn algorithm **ConstructTrie(D,P,T)** is shown in figure 1, Referring expression generation algorithm **MakeRefExpr(r,p,T,L) is** shown in figure 2, where **T** is a node pointer and **p** is pointer to parent of that node. Our algorithm **MakeRefExpr** returns set of attribute-values **L** to identify **r** in the domain. **[[$N_i$]]**= {**d** |**d**∈**D** and **d** is stored at node $\mathbf{N_i}$ where $\mathbf{N_i}$ is an i-th level node}. **Card(N)** is cardinality of set of objects in node **N**.

```
ConstructTrie(D, P, T) {
 If (D = ∅ ∨ P = ∅)
 Then Stop
 Else
   Create a node N at T
   Set ⟦ N ⟧ = D
   Extract front attribute A_i from list P
   P' = P – { A_i }
   For each value V_j of attribute A_i do
      Create Edge E_j with label V_j as T→Next_j
      D_j' = D ∩ ⟦ Val(E_j) ⟧
      ConstructTrie(D_j', P', T→Next_j)
   End For
 End If
}
```

Figure 1. Prefix Tree Generation Algorithm

```
MakeRefExpr(r, P, T, L) {
  If ( r ∉ ⟦ T→N ⟧ )
       Then L ← ∅
       Return L
  Else If ({r} = ⟦ T→N ⟧)
       L = L ∪ Val(P→E_j )
       Return L
  Else If (isLeaf (T) ∧ {r} ⊂ ⟦ N ⟧)
       Then L ← ∅
       Return L
  Else {
     If (Card(P→N) > Card (T→N))
        Then L = L ∪ Val(P→E_j )
     P = T
     For each outgoing edge T→ Next_j (E_j) do
       L' = MakeRefExpr(r, P,T→ Child_j, L)
          If (L' ≠ ∅)
          Then Return L'
     } }
```

Figure 2. Expression Generation Algorithm

The significant achievement is that incompleteness of previous algorithms can be tackled in this model in a straightforward way. For example, in case of vague descriptions (overlapping properties), Incremental and other algorithms are unable to find unambiguous description even if it exists but our prefix tree model takes into account hearer model

and generate description for identifying the target object. Besides, in case of Boolean, plural, context sensitive and relational description generation our model provides a simple and linguistically rich approach to GRE.

## 2 Evaluation Results

In Table 1 and 2 the evaluation results for Furniture and People data has been shown.

| TRIAL | DICE | MASI | A_ACCURACY | UNIQUENESS | MINIMALITY |
|---|---|---|---|---|---|
| s101t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s101t6 | 0.75 | 0.2 | 0.0 | 1.0 | 0.0 |
| s101t7 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s108t5 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s114t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s114t6 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s120t4 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s120t6 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s123t1 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| s131t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s169t1 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s169t3 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s18t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s191t3 | 0.66666 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s191t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s191t6 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s195t1 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s197t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s197t3 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s197t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s197t6 | 0.75 | 0.2 | 0.0 | 1.0 | 0.0 |
| s19t5 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s19t7 | 0.5 | 0.11111 | 0.0 | 1.0 | 0.0 |
| s20t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s20t6 | 0.66666 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s22t3 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s22t4 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s28t6 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s307t3 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s30t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s310t3 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s312t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s313t2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s315t2 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s31t4 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s31t6 | 0.28571 | 0.05555 | 0.0 | 1.0 | 0.0 |
| s320t2 | 0.66666 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s320t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s320t7 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s323t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s323t3 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s326t2 | 0.5 | 0.22222 | 0.0 | 1.0 | 0.0 |
| s326t6 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s329t5 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s331t6 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s346t2 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s367t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s367t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s373t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s373t6 | 0.66666 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s374t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s374t5 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s41t2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s41t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s48t4 | 0.66666 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s49t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s49t6 | 0.5 | 0.11111 | 0.0 | 1.0 | 0.0 |
| s50t6 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s53t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s53t2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s53t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s57t5 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s57t6 | 0.66666 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s59t3 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s60t2 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s62t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s62t5 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s62t6 | 0.5 | 0.11111 | 0.0 | 1.0 | 0.0 |
| s73t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s73t3 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s73t6 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s73t7 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s78t4 | 0.66666 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s78t7 | 0.57142 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s81t2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s81t4 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s84t1 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s84t6 | 0.66666 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s98t5 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s98t7 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |

Table1: Evaluation Result of Furniture data

| TRIAL | DICE | MASI | A_ACCURACY | UNIQUENESS | MINIMALITY |
|---|---|---|---|---|---|
| s101t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s102t23 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s106t23 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| s114t21 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s114t23 | 0.66667 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s123t25 | 0.28577 | 0.05556 | 0.0 | 1.0 | 0.0 |
| s131t21 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s131t22 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s132t24 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s132t25 | 0.28571 | 0.05556 | 0.0 | 1.0 | 0.0 |
| s169t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s177t22 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s181t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s18t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s191t22 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s191t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s192t25 | 0.66667 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s197t22 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s197t23 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s197t26 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s19t25 | 0.28571 | 0.05556 | 0.0 | 1.0 | 0.0 |
| s20t22 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s20t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s22t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s307t22 | 0.66667 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s30t25 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s310t25 | 0.57143 | 0.13333 | 0.0 | 1.0 | 0.0 |
| s312t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s315t21 | 0.66667 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s315t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s31t24 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s31t26 | 0.33333 | 0.06667 | 0.0 | 1.0 | 0.0 |
| s320t23 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s323t23 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s323t26 | 0.66667 | 0.16667 | 0.0 | 1.0 | 0.0 |
| s331t23 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s332t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s373t23 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s373t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s374t21 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s374t22 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s374t23 | 0.66667 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s40t26 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s41t23 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s48t21 | 0.75 | 0.2 | 0.0 | 1.0 | 0.0 |
| s49t21 | 0.28571 | 0.05556 | 0.0 | 1.0 | 0.0 |
| s49t22 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s49t23 | 0.8 | 0.44444 | 0.0 | 1.0 | 0.0 |
| s50t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s50t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s50t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s53t21 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s53t23 | 0.66667 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s53t26 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s57t23 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s57t24 | 0.4 | 0.08333 | 0.0 | 1.0 | 0.0 |
| s57t25 | 0.28571 | 0.05556 | 0.0 | 1.0 | 0.0 |
| s59t21 | 0.66667 | 0.33333 | 0.0 | 1.0 | 0.0 |
| s59t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s59t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s60t24 | 0.57143 | 0.26667 | 0.0 | 1.0 | 0.0 |
| s62t21 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| s69t25 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| s73t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s78t21 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |
| s78t25 | 0.88889 | 0.53333 | 0.0 | 1.0 | 0.0 |
| s81t24 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 |
| s81t25 | 0.85714 | 0.5 | 0.0 | 1.0 | 0.0 |

Table2: Evaluation Result of People data

## References


R. Dale and E. Reiter. 1995. Computational Interpretations of the Gricean Maxims in the generation of Referring Expressions. *Cognitive Science* (18): 233 –263

S. Paladhi and S. Bandyopadhyay. 2008. Generation of Referring Expression Using Prefix Tree Structure. Proceedings of IJCNLP: 697-702

van Deemter. 2002. Boolean Extensions of Incremental Algorithm. *Computational Linguistics* 28(1): 37-52