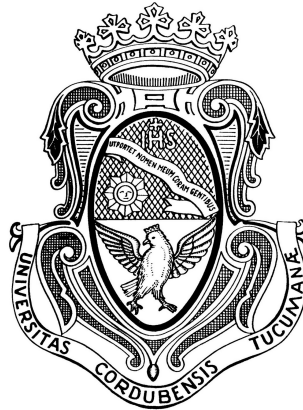


UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA



Trabajo final del Doctorado en Ciencias de la Computación

## **Generación de expresiones referenciales: el punto de vista lógico, el más humano**

Alumna: Ivana Romina Altamirano

Supervisora: Luciana benotti

CÓRDOBA ARGENTINA  
Diciembre, 2013



*Gracias Lu, Carlos . . .*



■ CLASIFICACIÓN DE BIBLIOTECA:

I.2.7

■ PALABRAS CLAVE:

Palabra Clave 1, *Palabra Clave 2*, Palabra Clave 3, Palabra Clave 4, *Palabra Clave 5*, *Palabra Clave 6*.

## RESUMEN

Se propone un algoritmo para la generación de expresiones referenciales (REs) que adapta la aproximación de [?, ?] para incluir sobre-especificación y probabilidades aprendidas a partir de corpus. Luego de introducir el algoritmo discutimos como las probabilidades requeridas pueden ser computadas para cualquier dominio para el cual existe un corpus de REs y como las probabilidades pueden ser ajustadas para nuevas escenas en el dominio usando aprendizaje automático. Nosotros ejemplificamos como computar las probabilidades sobre el corpus GRE3D7 de [?] y el TUNA corpus de [?]. El algoritmo resultante es capaz de generar diferentes expresiones referenciales para el mismo target con una frecuencia similar a aquella observada en el corpus. Nosotros empíricamente evaluamos el nuevo algoritmo sobre el corpus GR3D7, y mostramos que la distribución de probabilidad de las expresiones referenciales generadas por el algoritmo machean las encontradas en el corpus con alta precisión. Nosotros también comparamos nuestros resultados con los resultados the [?], los ganadores de la competencia NLG Challenge sobre selección de atributos para generación de expresiones referenciales (ASGRE). También proveemos un análisis de error y conclusiones para ambos corpus.

## ABSTRACT

We propose an algorithm for the generation of referring expressions (REs) that adapts the approach of arec2:2008:Areces,arec:usin11 to include overspecification and probabilities learned from corpora. After introducing the algorithm, we discuss how probabilities required as input can be computed for any given domain for which a suitable corpus of REs is available, and how the probabilities can be adjusted for new scenes in the domain using a machine learning approach. We exemplify how to compute probabilities over the GRE3D7 corpus of viet:gene11 and the TUNA corpus of gatt-balz-kow:2008:ENLG. The resulting algorithm is able to generate different referring expressions for the same target with a frequency similar to that observed in corpora. We empirically evaluate the new algorithm over the GRE3D7 corpus, and show that the probability distribution of the generated referring expressions matches the one found in the corpus with high accuracy. We also compare our results with the results of graph08 the winners of the NLG Challenge on Attribute Selection for Generating Referring Expressions (ASGRE). We provide error analysis and conclusion for both corpus.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Descripción del problema . . . . .	2
1.2. Contribuciones . . . . .	2
1.3. Mapa de la tesis . . . . .	2
<b>2. Selección de contenido de expresiones referenciales</b>	<b>3</b>
2.1. Generación automática de expresiones referenciales . . . . .	4
2.2. Algoritmos de REG . . . . .	4
2.2.1. Incremental . . . . .	4
2.2.2. Graph . . . . .	4
2.2.3. Bisimulation . . . . .	4
2.3. Aproximaciones empíricas a la solución de REG . . . . .	4
2.3.1. Corpus existente . . . . .	4
2.3.2. Jette y otros trabajos empíricos . . . . .	4
2.3.3. Métricas de evaluación/comparación con corpus . . . . .	4
2.4. Optimalidad de expresiones referenciales . . . . .	4
2.4.1. Grice... . . . .	4
2.4.2. Paraboni? . . . . .	4
2.4.3. Naturalidad Dale, Jette . . . . .	4
<b>3. Decisiones metodológicas de corpus y evaluación</b>	<b>5</b>
3.1. Corpus problemas de corpus existente y/o de anotación . . . . .	6
3.2. Comparando la salida del sistema con ER hechas por humanos . . . . .	6
3.3. Problemas en la evaluación de GRE . . . . .	6
3.4. Corpus seleccionados . . . . .	6
<b>4. Recolección y análisis del corpus ZOOM</b>	<b>7</b>
4.1. Características del corpus . . . . .	8
4.2. Método de recolección del corpus . . . . .	8
4.3. Anotación del corpus . . . . .	8
<b>5. Nuestra propuesta</b>	<b>9</b>
5.1. Agregando probabilidades a un algoritmo existente . . . . .	10
5.2. Aprendizaje automático . . . . .	10
5.3. Agregando sobreespecificación . . . . .	10
<b>6. Evaluación de nuestra propuesta</b>	<b>11</b>
<b>7. Conclusiones</b>	<b>13</b>





## Capítulo 1

# Introducción

## 1.1. Descripción del problema

En lingüística una expresión referencial (RE) es una expresión que identifica unívocamente a un objeto de para un interlocutor, desde un conjunto de posibles distractores. Por ejemplo si nosotros queremos identificar a un cierto animal  $d$  de un conjunto de mascotas, la expresión “el perr” será ER si  $d$  es el único perro en el conjunto, y si nosotros estamos seguros que nuestro interlocutor identificará a  $d$  como un perro.

## 1.2. Contribuciones

- Un algoritmo para la generación de expresiones referenciales independiente de dominio.
- 

## 1.3. Mapa de la tesis

Esta tesis se divide en 6 capítulos, en el primer capítulo ...

## Capítulo 2

# Selección de contenido de expresiones referenciales

## **2.1.    Generación automática de expresiones referenciales**

## **2.2.    Algoritmos de REG**

### **2.2.1.    Incremental**

### **2.2.2.    Graph**

### **2.2.3.    Bisimulation**

## **2.3.    Aproximaciones empíricas a la solución de REG**

### **2.3.1.    Corpus existente**

### **2.3.2.    Jette y otros trabajos empíricos**

### **2.3.3.    Métricas de evaluación/comparación con corpus**

## **2.4.    Optimalidad de expresiones referenciales**

### **2.4.1.    Grice...**

### **2.4.2.    Paraboni?**

### **2.4.3.    Naturalidad Dale, Jette**

## Capítulo 3

# Decisiones metodológicas de corpus y evaluación

- 3.1. Corpus problemas de corpus existente y/o de anotación
- 3.2. Comparando la salida del sistema con ER hechas por humanos
- 3.3. Problemas en la evaluación de GRE
- 3.4. Corpus seleccionados

## Capítulo 4

# Recolección y análisis del corpus ZOOM

- 4.1. Características del corpus
- 4.2. Método de recolección del corpus
- 4.3. Anotación del corpus



## Capítulo 5

# Nuestra propuesta

- 5.1. Agregando probabilidades a un algoritmo existente
- 5.2. Aprendizaje automático
- 5.3. Agregando sobreespecificación

## Capítulo 6

# Evaluación de nuestra propuesta



## Capítulo 7

## Conclusiones