



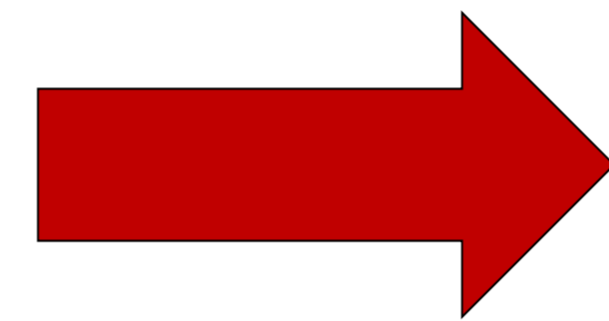
# The Prevalence of Descriptive Referring Expressions in News and Narrative

**Raquel Hervás**  
Facultad de Informática,  
Universidad Complutense de Madrid  
Spain  
raquelhb@fdi.ucm.es  
<http://nil.fdi.ucm.es>

**Mark Finlayson**  
Computer Science and  
Artificial Intelligence Laboratory,  
Massachusetts Institute of Technology  
USA  
markaf@mit.edu  
<http://www.mit.edu/~markaf>

What is a *descriptive* referring expression?

Once upon a time there was a man, who had three daughters. They lived in a house and their dresses were made of fabric.



Once upon a time there was a rich merchant, who had three daughters. They lived in a very fine house and their gowns were made of the richest fabric sewn with jewels.

Text with only distinctive referring expressions

Real text with *descriptive* referring expressions

→ Current approaches to referring expression generation (REG) focus almost exclusively on how to generate distinctive referring expressions.

## Question:

How prevalent are descriptive referring expressions?  
(secondary question: is descriptiveness found mostly in modifiers?)

If descriptive referring expressions are relatively prevalent, this suggests we need to rethink a lot of REG algorithms

**Approach:** A corpus study of two different genres

## Constituents: Nuclei vs. Modifiers

[Nuclei]: portions that perform the 'core' referring function

(Modifiers): portions that can be varied, syntactically speaking, independently of the nuclei

- (The) [roof] (of the house) collapsed.
- (The) (city of) [Boston] is beautiful

## Functions: Distinctive vs. Descriptive

**Descriptive<sub>c</sub>:** If a constituent of a referring expression is (a) unnecessary for identifying the referent or (b) expressed using unusual or ostentatious word choice.

**Distinctive<sub>f</sub>:** Any constituent that is not descriptive.

- John had a black dog, and Mary a white one. (The)<sub>f</sub>-(black)<sub>f</sub>-(dog)<sub>f</sub> was a barker.
- John liked (the)<sub>f</sub>-(white)<sub>c</sub>-(dog)<sub>f</sub>. Mary liked it too.

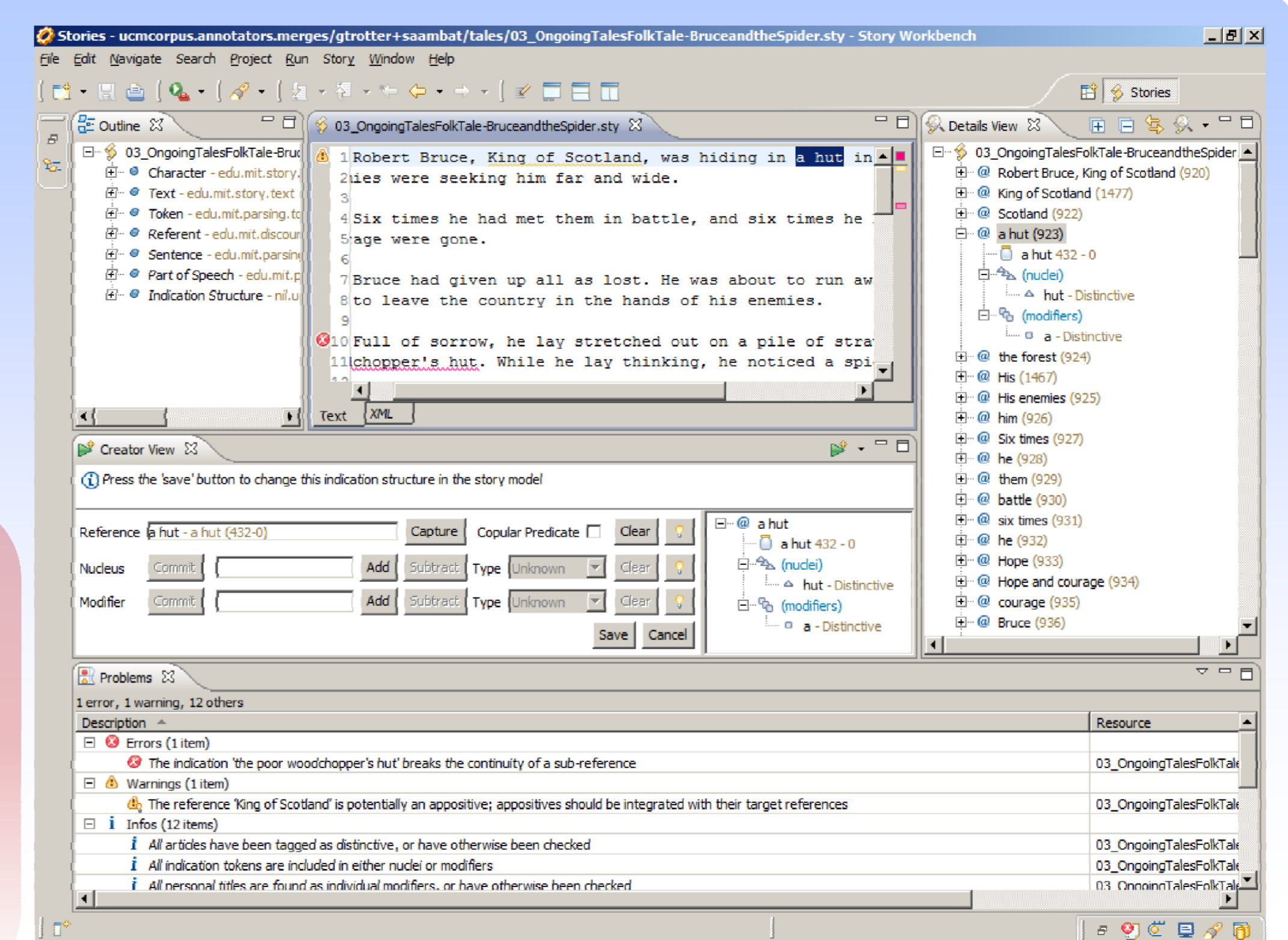
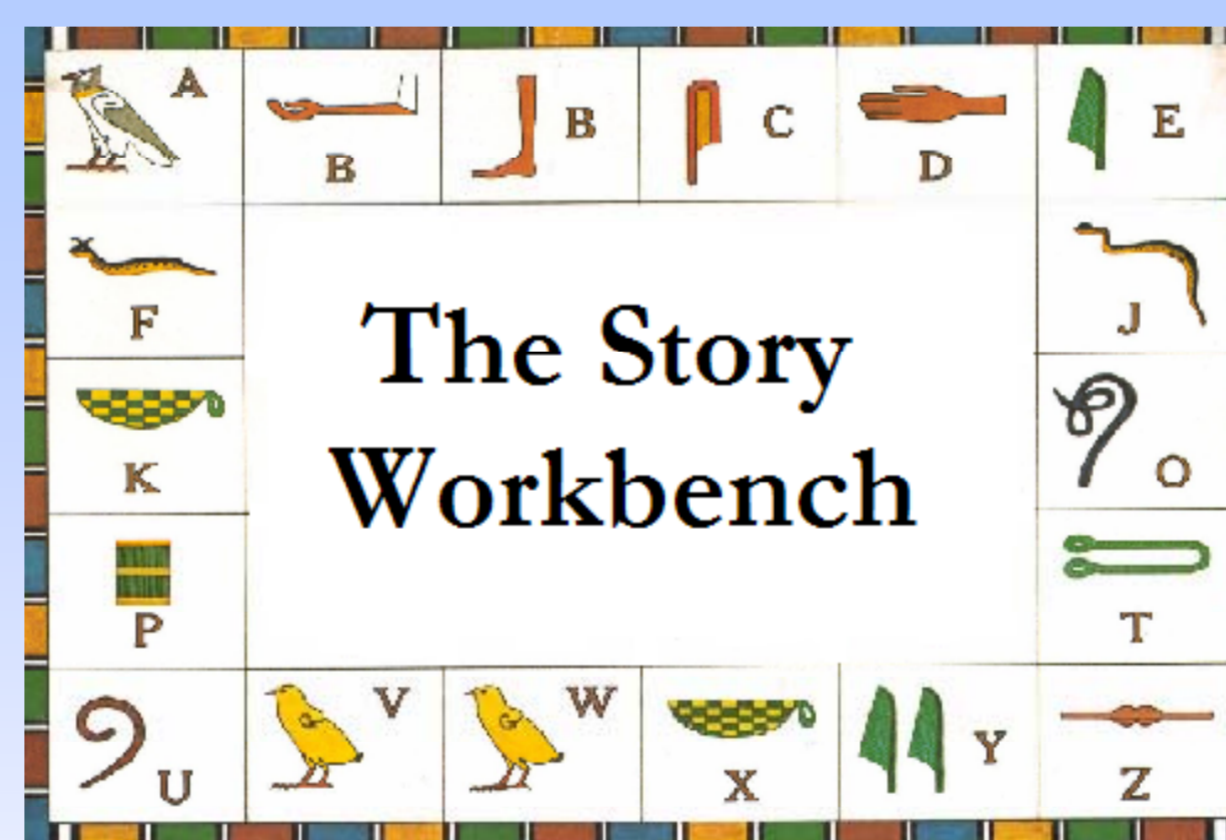
DEFINITIONS

ANNOTATION

Two Annotators, native English speakers

Annotation Guide at (Finlayson & Hervas 2010a)

Annotation performed using the Story Workbench annotation tool (Finlayson 2008)



UMIREC Corpus (UCM/MIT Indications, Referring Expressions, and Coreference)

- 24,422 words
- 2 genres (news articles from WSJ Penn Treebank sections, variety of western european folktales)
- 62 texts total (32 news, 30 folktales)
- Double-annotated & adjudicated
- Referring expression locations, constituents & functions, co-reference relationships

Download from <http://hdl.handle.net/1721.1/54766>

RESULTS

## Corpus

	Tales	Articles	Total
Texts	30	32	62
Words	12,050	12,372	24,422
Sentences	904	571	1475
Referring Expressions	3,681	3,526	7,207

Corpus at (Finlayson & Hervas 2010b)

## Agreement

	Tales	Articles	Total
Referring Expressions (F1)	1	0.99	0.99
Constituents (F1)	0.99	0.98	0.98
Nucleus / Modifier Label (K)	0.97	0.95	0.96
Constituent Function (K)	0.61	0.48	0.54
Referring Expression Function (K)	0.65	0.54	0.59

F1 = Balanced F-measure, K = kappa measure for 2 tags

## Counts

	Tales	Articles	Total
Distinctive Nuclei	95%	97%	96%
Descriptive Nuclei	5%	3%	4%
Distinctive Modifiers	78%	81%	80%
Descriptive Modifiers	22%	19%	20%
Distinctive Referring Expressions	83%	81%	82%
Descriptive Referring Expressions	17%	19%	18%

Nearly one-fifth of referring expressions are descriptive!

in news & narrative