

Abstract

Definiteness expresses a constellation of semantic, pragmatic, and discourse properties (the communicative functions) of an NP. Our supervised classifier for English NPs uses lexical, morphological, and syntactic features to predict the communicative functions in terms of a language-universal classification scheme and establishes strong baselines for future work. Additionally, analysis of the features and learned parameters in the model provides insight into the grammaticalization of definiteness in English, not all of which is obvious a priori.

Classification Model

We use an in-house implementation of a multiclass logistic regression classifier with L2-regularization [***AB: Do we want to say why we used L2, why log regression? Also how do I make the following equatoin look better? I don't want the first part to be in the center, but more to the left.***].

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} -\lambda ||\boldsymbol{\theta}||_2^2 + \sum_{\langle x,y \rangle \in \mathcal{D}} \log \frac{\exp \boldsymbol{\theta}^\top \mathbf{f}(x,y)}{\sum_{y' \in \mathcal{Y}} \exp \left(\boldsymbol{\theta}^\top \mathbf{f}(x,y') + \kappa \operatorname{cost}(y,y') \right)}$$

Features

Words of Interest Head of the NP, its dependents, its governor (external to NP), its first ancestor verb — token, lemma, POS tag, dependency relation, a binary indicator of plurality on the head N, first_dependent, last_dependent, auxiliaries of the first ancestral verb, first ancestral verb with a negative particle as dependent.

Structural — path length to the root, path length to the first ancestral verb, number of dependents, number of dependency relations that link non-neighbors.

Positional — token length of the NP, NP's location in the sentence (first or second half), the first ancestral verb's position relative to the head (left or right), POS & lemma of the left and the right neighbors of the head, governor, and the first ancestral verb.

Above features of NPs in Following NP-NP relation Types immediate parent, immediate child, immediate precedent, immediate successor, the nearest preceding coreferent mention.

The language independent scheme for Communicative Functions of Definiteness

Nonanaphora		Anaphora & Miscellaneous	
Nonanaphora [-A,-B]	999	Anaphora [+A]	1574
- Unique [+U]	287	- Basic_Anaphora [-B,+F]	795
* Unique_Hearer_Old [+F,-G,+S]	251	*Same_Head	556
Unique_Physical_Copresence [+R]	13	*Different_Head	329
Unique_Larger_Situation [+R]	237	- Extended_Anaphora [+B]	779
Unique_Predicative_Identity [+P]	1	*Bridging_Nominal [-G,+R,+S]	43
*Unique_Hearer_New [-F]	36	*Bridging_Event [+R,+S]	10
- Nonunique [-U]	581	*Bridging_Restrictive_Modifier [-G,+S]	614
* Nonunique_Hearer_Old [+F]	169	*Bridging_Subtype_Instance [-G]	0
Nonunique_Physical_Copresence [-G,+R,+S]	39	*Bridging_Other_Context [+F]	112
Nonunique_Larger_Situation [-G,+R,+S]	117		
Nonunique_Predicative_Identity [+P]	13	Miscellaneous [-R]	732
*Nonunique_Hearer_New_Spec [-F,-G,+R,+S]	231	- Pleonastic [-B,-P]	53
*Nonunique_Nonspec [-G,-S]	181	- Quantified	248
- Generic [+G,-R]	131	- Predicative_Equative_Role [-B,+P]	58
*Generic_Kind_Level	0	- Part_Of_Noncompositional_MWE	100
*Generic_Individual_Level	131	- Measure_Nonreferential	125
		- Other_Nonreferential	148

Examples for Communicative Functions

CFD Label	Example
Unique_Physical_Copresence Unique_Larger_Situation	John here is an investment banker. In the days since Hillary Clinton unburdened herself in an interview with The Atlantic's Jeffrey Goldberg ...
Unique_Predicative_Identity Unique_Hearer_New Nonunique_Physical_Copresence Nonunique_Larger_Situation Nonunique_Predicative_Identity Nonunique_Hearer_New_Specific Nonunique_Nonspec Generic_Kind_Level Generic_Individual_Level Basic_Same_Head Basic_Different_Head Extended_Bridging_Nominal Extended_Bridging_Event	Clark Kent is Superman . a restaurant chain named Shoney's The podium is too high. the chair (at a conference) / today He is the manager . I am looking for a nurse . Her name is Sara. I am looking for a nurse [any nurse would do]. Dinosaurs are extinct. Cats have fur. I'm going to tell you <u>a quick story</u> . It's a true story . I adopted <u>a cat</u> this weekend. The animal is so cute. I looked at <u>an apartment</u> yesterday. The kitchen was really large. My friend's son <u>got married</u> this weekend. The bride looked beautiful.
Extended_Bridging_Restrictive_Modifier Extended_Subtype_Instance Extended_Other_Context	the house <u>next door</u> / <u>John's</u> daughter I collect <u>coins</u> . I have a 1943 steel penny . I want to focus on what many of you have said you would like me to elaborate on. What can you do about the climate crisis ?
Pleonastic Quantified Predicative_Equative_Role Part_of_Noncompositional_MWE Measure_Nonreferential Other_Nonreferential	It is raining . All the people / no motorcade He's a teacher . / This is an opportunity . Ole' Charlie kicked the bucket today. hours later / miles away global warming / concern / the topic of energy

Accuracy

Condition	# Params	ExactMatch(%)	SoftMatch(%)
Majority baseline	—	12.1	47.8
Log			
+ attributes	473,064	38.7	77.1
+ labels	413,931	40.8	73.6
+ attributes, labels	926,417	43.7	78.2
Random forest	20,363	49.7	77.5
Leaf label	N	F1	
BRIDGING_RESTRICTIVE_MODIFIER	552	68	
SAME_HEAD	452	41	
DIFFERENT_HEAD	271	32	
PLEONASTIC	44	88	
[***AB: will add all teh other labels a bit later.***]			

Analysis

Confirmation of known facts ...

Good hypotheses to test ...

Baffling cases ...

[***AB: to be added soon***]

Acknowledgements

This research was sponsored by a grant from the U.S. Army Research Lab and the U.S. Army Research Office.