# Adapting Coreference Resolution for Narrative Processing

**Quynh Ngoc Thi Do**[1], **Steven Bethard**[2], **Marie-Francine Moens**[1]
[1]Katholieke Universiteit Leuven, Belgium
[2]University of Alabama at Birmingham, United States
quynhngocthi.do@cs.kuleuven.be
bethard@cis.uab.edu
sien.moens@cs.kuleuven.be

## Abstract

Domain adaptation is a challenge for supervised NLP systems because of expensive and time-consuming manual annotated resources. We present a novel method to adapt a supervised coreference resolution system trained on newswire to short narrative stories without retraining the system. The idea is to perform inference via an Integer Linear Programming (ILP) formulation with the features of narratives adopted as soft constraints. When testing on the UMIREC[1] and N2[2] corpora with the-state-of-the-art Berkeley coreference resolution system trained on OntoNotes[3], our inference substantially outperforms the original inference on the CoNLL 2011 metric.

## 1 Introduction

Coreference resolution is the task of partitioning the set of mentions of discourse referents in a text into classes (or 'chains') corresponding to those referents (Stede, 2011). To solve the problem, contextual and grammatical clues, as well as semantic information and world knowledge are necessary for either learning-based (Bengtson and Roth, 2008; Stoyanov et al., 2010; Haghighi and Klein, 2010) or rule-based (Haghighi and Klein, 2009; Lee et al., 2011) coreference systems. These systems draw on diverse information sources and complex heuristics to resolve pronouns, model discourse, determine anaphoricity, and identify semantically compatible mentions. However, this leads to systems with many hetorogenous parts that can be difficult to interpret or modify.

Durrett and Klein (2013) propose a learning-based, mention-synchronous coreference system to tackle the various aspects of coreference by using the simplest possible set of features. Its advantage is that the system can both implicitly model important linguistic effects and capture other patterns in the data that are not easily teased out by hand. With a simple set of features including head/first/last words, preceding/following words, length, exact string match, head match, sentence/mention distance, gender, number etc. and an efficient training using conditional log-likelihood augmented with a parameterized loss function optimization they report state-of-the-art results on CoNLL 2011 data.

But while CoNLL 2011 training data (OntoNotes) includes a few different source domains (newswire, weblogs, etc.), we witness significant drops in performance when systems trained on CoNLL 2011 are applied to new target domains such as narratives. Some linguistic effects and patterns that are very important for the target domain were never seen in the source domain on which the model was trained. In such cases, when adapting a coreference system to a new domain, it is necessary to incorporate these more complex linguistic features and patterns into the model.

We propose a novel method to adopt the target domain's features to a supervised coreference system without retraining the model. We present a case of transferring the system of (Durrett and Klein, 2013), which is trained on OntoNotes, to short narrative stories. The idea is to perform inference via a linear programming formulation with the features of narratives adopted as soft constraints. Since the new features are incorporated only into the linear program, there is no need to retrain the original model. Our formulation models three phenomena that are important for short narrative stories: local discourse coherence, which we model via centering theory constraints, speaker-listener relations, which we model via direct speech act constraints, and character-naming, which we model via definite noun phrase and exact match constraints.

[1] http://dspace.mit.edu/handle/1721.1/57507
[2] http://dspace.mit.edu/handle/1721.1/85893
[3] https://catalog.ldc.upenn.edu/LDC2011T03

*Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2262–2267,
Lisbon, Portugal, 17-21 September 2015.

We also suggest a method to compute back pointers (as defined in Durrett and Klein (2013)) globally.

## 2 Berkeley coreference system

Given $N$ mentions $m_1, ..., m_N$ from a document $x$, each $m_i$ has an associated random variable $a_i$ taking values in the set of $\{1, ..., i-1, NEW\}$. This variable specifies $m_i$'s selected antecedent or indicates that it begins a new coreference chain. We call $a_i$ the *back pointer* of $m_i$. A setting of all the back pointers, denoted by $a = (a_1, ..., a_n)$, implies an unique set of coreference chains that serve as the system output.

A log-linear model of the conditional distribution $P(a|x) \propto \exp \sum_{i=1}^{n} \mathbf{f}(i, a_i, x)$ is used, where $\mathbf{f}(i, a_i, x)$ is a feature function that examines the coreference decision $a_i$ for $m_i$ with document context $x$. If $a_i = NEW$, the features indicate the suitability of the given mention to be anaphoric or not; when $a_i = j$ for some $j$, the features express aspects of the pairwise linkage, and examine relevant attributes of the anaphor $i$ or the antecedent $j$. During training, the model is optimized with a parameterized loss function. The inference is simple and efficient: because $\log P(a|x)$ decomposes linearly over mentions, $a_i = \arg\max_{a_i} P(a_i|x)$ (Durrett and Klein, 2013).

## 3 Computing back pointers globally

A drawback of computing each $a_i$ locally is that the system does not take into account constraints from mentions outside of the (mention, antecedent) pairs. For example, given three mentions $m_1, m_2, m_3$, if the system predicts that $a_2 = 1$ and $a_3 = 2$ (i.e., that $m_2$'s antecedent is $m_1$ and $m_3$'s antecedent is $m_2$), then $m_3$ will be automatically inferred as coreferent with $m_1$. But if there is a clear clue that $m_1$ and $m_3$ are not coreferent, leveraging this clue could help avoid the error of linking $m_3$ to $m_2$.

In this work, we perform inference via an ILP formulation which allows new linguistic features and patterns over mentions – not only (mention, antecedent) pairs – that were not part of training the original model to be adopted as constraints of the ILP problem.

Let $\mathbf{U}$ be the set of binary indicator variables corresponding to the values assigned to the back pointers. Specifically, $u_{ij} = 1$ iff $a_i = j$ and $u_{ii} = 1$ iff $a_i = NEW$.

$\mathbf{C}$ is the set of $K$ binary constraint indicator variables indicating if linguistic constraints are violated. Specifically, $c_{k,i,j} = 1$ iff the linguistic constraint $C_k$ is violated for the back pointer $u_{ij}$. Each $C_k$ is associated with a penalty score $\rho_k$.

We aim to maximize the objective function:

$$\sum_{i=1}^{N} \sum_{j=1}^{i} u_{ij} P(a_i = j|x) - \sum_{k=1}^{K} \rho_k c_{k,i,j} \quad (1)$$

Subject to:

$$\forall i : \sum_{j=1}^{i} u_{ij} = 1$$

To incorporate coreference constraints, we introduce $\mathbf{V}$, a set of binary variables indicating if two mentions are in the same coreference chain. For each pair of $j < i$, a variable $v_{ij}$ is added to the ILP model, where $v_{ij} = 1$ iff $m_i$ and $m_j$ are in the same chain. The definition of $v_{ij}$ in terms of $u_{ij}$ is encoded as the following ILP constraints:

$$\begin{aligned} \forall j < i : u_{ii} + v_{ij} &\leq 1 \\ \forall j < i : u_{ij} - v_{ij} &\leq 0 \\ \forall k < j < i : u_{ij} + v_{jk} - v_{ik} &\leq 1 \\ \forall k < j < i : u_{ij} - v_{jk} + v_{ik} &\leq 1 \\ \forall j < k < i : u_{ij} + v_{kj} - v_{ik} &\leq 1 \\ \forall j < k < i : u_{ij} - v_{kj} + v_{ik} &\leq 1 \end{aligned}$$

For long texts, to reduce the complexity of the ILP problem, we set a threshold, $windows_v$, so that $v_{ij}$ is only available if $i - windows_v \leq j$.

The framework of $\mathbf{V}$ variables allows coreference constraints to be adopted easily by any coreference resolution system that provides scores for each possible back pointer value. For example, consider the Stanford exact string match sieve, which "requires an exact string match between a mention and its antecedent" (Lee et al., 2011). If we want to encourage such matches, for each pair $j < i$ where the two nominal mentions $m_i$ and $m_j$ have an exact string match, we would introduce a constraint indicator variable $c_{exact,i,j}$ and add the constraint $v_{ij} + c_{exact,i,j} = 1$ to the ILP model. The result would be that when the exact match constraint is violated and some $v_{ij} = 0$, ILP would force the corresponding $c_{exact,i,j} = 1$ and the objective function would be reduced by $\rho_{exact}$.

ILP has been used previously to enforce global consistency in coreference resolution (Finkel and Manning, 2008; Denis and Baldridge, 2007; Barzilay and Lapata, 2006). These models were designed for an all-pairs classification approach to coreference resolution, and are not directly applicable to the back pointer approach of (Durrett and Klein, 2013). But the back pointer approach allows features to be expressed more naturally using local context, rather than requiring, for example, judgments of whether two pronouns separated by many paragraphs are coreferent. Moreover, our ILP formulation is the only one to consider the problem of adapting to another domain and incorporating new features without retraining the original model.

## 4 Centering theory constraints

Pronouns, in particular, have a huge effect on information flow across sentences. Since they are almost void of meaning (only signal gender and number of the antecedent), the discourse referent to be picked up must be particularly salient, so that it can be readily identified by the reader (Stede, 2011). The discourse center hypothesis (Hudson-D'Zmura, 1988) states that at any point in discourse understanding, there is one single entity that is the most salient discourse referent at that point. This referent is called the center. Centering theory is a key element of the discourse center hypothesis used in anaphora resolution (Grosz et al., 1995). Beaver (2004) reformulates the centering theory in terms of Optimality Theory (Prince and Smolensky, 2004). Six ranked constraints – Agree, Disjoint, ProTop, FamDef, Cohere and Align – are used to make anaphora decisions. We adopt four of these constraints in our ILP model as follows:

**Disjoint** "Co-arguments of a predicate[4] are disjoint." For each $j < i$ such that $m_i$ and $m_j$ are subject and object arguments of a non-reflexive predicate, we introduce a constraint indicator variable $c_{disjoint,i,j}$, and add the ILP constraint $v_{ij} - c_{disjoint,i,j} = 0$.

**ProTop** "The *topic* of a sentence which is the entity referred to in both the current and the previous sentence, is pronominalized." If a sentence contains pronouns then at least one of its pronouns is coreferent with a mention in the previous sentence. For each sentence $t$ containing pronouns, we introduce a constraint indicator variable $c_{protop,t,t-1}$, and add the ILP constraints:

$$\forall i \in \mathbf{P}_t, \forall j \in \mathbf{M}_{t-1} : v_{ij} + c_{protop,t,t-1} \leq 1$$
$$c_{protop,t,t-1} + \sum_{i \in \mathbf{P}_t} \sum_{j \in \mathbf{M}_{t-1}} v_{ij} \geq 1$$

$\mathbf{P}_t$ is the set of all pronouns in sentence $t$. $\mathbf{M}_{t-1}$ is the set of all mentions in sentence $t-1$ [5].

**FamDef** "No new information about the referent is provided by the definite." We consider only pronouns here, though the original FamDef also includes definite descriptions and proper names (Beaver, 2004). For each pronoun $m_i$, we introduce a constraint indicator variable $c_{famdef,i,i}$ and add the ILP constraint $u_{ii} - c_{famdef,i,i} = 0$.

**Align** "The topic is in subject position." More specifically, the topic of a sentence is pronominalized and prefers the subject position over other positions. For each sentence containing only one pronoun $m_i$, if the previous sentence has only one verbal semantic frame and $m_j$ is its subject, we introduce a constraint indicator variable $c_{align,i,j}$, and add the ILP constraint $v_{ij} + c_{align,i,j} = 1$.

Note: The ProTop, FamDef and Align constraints are not applied to sentences containing quotations.

## 5 Direct speech constraints

Direct speech acts (with quotation marks) are detected and attached to the closest verbal communication semantic frames. For each direct speech act $q_t$, we call the mentions $m_{st}, m_{ot}$ the speaker and listener of $q_t$ if they play the subject and object roles respectively in the semantic frame of $q_t$. We detect the set of subject pronouns[6] inside the quote marks of $q_t$ and name it $\mathbf{S}_t$. The set of all mentions that refer to the speaker of $q_t$ is $\textsc{Speaker}_t = \{m_{st}\} \cup \mathbf{S_t}$. For each $(m_i, m_j) \in \textsc{Speaker}_t \times \textsc{Speaker}_t$ with $i > j$, we introduce a constraint indicator variable $c_{subject,i,j}$, and add the ILP constraint $v_{ij} + c_{subject,i,j} = 1$.

Similarly, $\mathbf{O_t}$ is the set of object pronouns[7] inside the quote marks of $q_t$. The set of all mentions that refer to the listener of $q_t$ is $\textsc{Listener}_t = \{m_{ot}\} \cup \mathbf{O_t}$. For each pair of mentions $(m_i, m_j) \in \textsc{Listener}_t \times \textsc{Listener}_t$ with $i > j$, we introduce a constraint indicator variable $c_{object,i,j}$, and add the constraint $v_{ij} + c_{object,i,j} = 1$.

If a conversation is detected (a sequence of "question" and "answer" semantic frames), the subject of the "question" is coreferent with the object

[4] A word that evokes a semantic frame (event) in a sentence.

[5] We can relax the constraint by replacing $\mathbf{M}_{t-1}$ with $\mathbf{M}_{t-1} \cup \mathbf{M}_{t-2} \cup \mathbf{M}_{t-3}$

[6] ("I", "me", "my", "mine", "myself") if $m_{st}$ is singular or ("we", "us", "our", "ourself") if $m_{st}$ is plural

[7] ("you", "your", "yours", "yourself")

| Method | UMIREC (Tales) | | | | N2 (Hadith) | | | |
|---|---|---|---|---|---|---|---|---|
| | MUC | BCUB | CEAFE | AVG | MUC | BCUB | CEAFE | AVG |
| ILPI with gold mentions | **84.16** | **65.65** | **50.47** | **66.76** | **80.47** | **65.53** | **54.06** | **66.69** |
| BER with gold mentions | 80.58 | 60.96 | 42.48 | 61.34 | 76.28 | 62.66 | 45.48 | 61.47 |
| ILPI with predicted mentions | **73.32** | **59.18** | **37.54** | **56.68** | **66.13** | **62.55** | **40.51** | **56.40** |
| BER with predicted mentions | 72.71 | 58.12 | 35.76 | 55.53 | 64.87 | 59.60 | 37.96 | 54.14 |

Table 1: ILPI and BER inference results on UMIREC (Tales) and N2 (Hadith) data.

of the "answer" and vice versa. For each pair of direct speech acts ($q_t$, $q_{t+1}$) that is a ("question", "answer") pair, for each pair of mentions $(m_i, m_j) \in \{\textsc{Listener}_{t+1} \times \textsc{Speaker}_t\} \cup \{\textsc{Speaker}_{t+1} \times \textsc{Listener}_t\}$, we introduce a constraint indicator variable $c_{conversation,i,j}$ and add the ILP constraint $v_{ij} + c_{conversation,i,j} = 1$.

## 6 Definite noun phrase and exact match constraints

In short narrative stories, characters are frequently named via proper names, pronouns or definite noun phrases (Toolan, 2009). Character names are repeated regularly over the whole stories. A character is often first presented as an indefinite noun phrase (such as "a woman"), then later as a definite noun phrase (such as "the woman"). In this work we introduce the definite noun phrase constraint: For each pair $j < i$, if $m_j$ is the indefinite form and $m_i$ is the definite form of the same noun phrase, to enforce that $m_i$ and $m_j$ are coreferent, we introduce a constraint indicator variable $c_{name,i,j}$, and add the ILP constraint $v_{ij} + c_{name,i,j} = 1$. To boost the identification of characters in the stories, the definite noun phrase constraint is used together with the exact match constraint (See Section 3) applied to noun phrases and proper nouns.

## 7 Experiment

We test our model on 30 English folktales from the UCM/MIT Indications, Referring Expressions, and Coreference (UMIREC) Corpus v1.1 (Finlayson and Hervs, 2010), and 64 text stories from the Hadith section of the Narrative Networks (N2) Corpus (Finlayson et al., 2014). The texts are preprocessed using the Stanford sentence splitter (Manning et al., 2014)[8] and the Berkeley coreference system's preprocessor. The Berkeley coreference system is trained on OntoNotes (newswire, broadcast news/conversation, and web texts). We use Gurobi[9] to solve our ILP problem, and the Lund semantic role labeler (Björkelund et al., 2009) to detect semantic frames. Note that in our implementation, "subject" and "object" used in Section 4 and Section 5 refer to "subject role" and "object role" of the semantic frames respectively. We use a separate section of the N2 corpus, the Inspire story texts, as the held-out validation set used for parameter tuning, resulting in $window_v = 40$, $\rho_{subject} = \rho_{object} = \rho_{conversation} = \rho_{definite} = \rho_{exact} = \rho_{disjoint} = 1$, $\rho_{protop} = 0.2$, $\rho_{famdef} = 0.2$, $\rho_{align} = 0.1$.

We compare our ILP inference (ILPI) to the standard Berkeley coreference system (BER) with both gold and predicted mentions. Table 1 shows that our inference improves the MUC, BCUB and CEAFE scores on both datasets, especially when using gold mentions[10]. The average ILP running times are 42.37s per UMIREC document and 22.7s per N2 document on a Core I7 2.3 GHz quad-core computer. Table 2 shows the effects of each constraint type when used alone. Surprisingly, the simplest constraint type (definite & exact match constraints) gives us the best improvement especially in terms of CEAFE score. This may be because definite & exact match constraint links mentions in the whole document, while the centering theory and direct speech act constraints are more local. And since short narrative stories often have a small set of characters (usually represented by definite noun phrases or proper nouns), when these characters are linked correctly, the coreference resolution result is improved considerably.

## 8 Discussion

Our method provides a promising solution when retraining a system is impossible or difficult. However, it may raise a question of the computing cost

[8] If two direct speech acts enclosed in quotation marks are adjacent and one is placed at the end of a sentence, we separate them into two different sentences.

[9] http://www.gurobi.com/

[10] Using gold mentions, our method also improves the score on the CoNLL 2011 test set by +1.11% (AVG: 72.46).

| Constraint | MUC | BCUB | CEAFE | AVG |
|---|---|---|---|---|
| Centering theory | 81.15 | 61.80 | 43.01 | 61.99 |
| Direct speech | 81.26 | 62.74 | 42.93 | 62.31 |
| Definite & Exact | 83.09 | 62.85 | 49.60 | 65.18 |

Table 2: Effects of different constraints on ILP inference on UMIREC (Tales) with gold mentions.

for tuning penalty scores especially with the large number of constraints. In such these cases, dividing the constraints into different groups where all constraints in the same group have the same penalty score may help to limit the number of scores that need to be tuned. In our case study, the system is not very sensitive to the values of the penalty parameters. If we set all the penalty scores to 1, the final AVG results on UMIREC and N2 corpus are 66.05 and 66.68 respectively[11]. Those scores are a bit less than the scores obtained after tuning parameters but still higher than the results obtained without ILP. Regardless, it's true that the proposed ILP approach is not necessarily less costly in some settings, but it can be applied to any coreference system that provides back pointers, not just the Berkeley one.

Instead of adopting features of the target domain as soft constraints as in our method, one may consider to use them as linguistic features and retrain the model. A simple domain adaptation approach by augmenting the feature space (Daumé et al., 2010) based on a limited set of annotated data in the target domain might be an alternative solution. But note that our approach does not use any annotated data of the target domain. Also, an unsupervised system as (Lee et al., 2011) might encode the target domain features (exact match noun phrases, direct speech act) as sieves (**hard**), but with the **soft** constraints, our system is more flexible when making global decisions.

Our approach can be applied to another target domain, such as bio-medical domain where we have entities and a list of acronyms in texts. Constraining the entities with their acronyms might help to improve the coreference resolution for bio-medical texts.

## 9 Conclusion

We have proposed a novel approach to adapt a supervised coreference resolution system trained on newswire domain to short narrative stories without retraining the system by modeling the inference as an ILP problem with the features of narratives adopted as soft constraints. Three phenomena that are important for short narrative stories: local discourse coherence, speaker-listener relations, and character-naming are modeled via centering theory, direct speech act and definite noun phrase & exact match constraints. We obtain promising results when transferring the Berkeley coreference resolution trained on OntoNotes to UMIREC (Tales) and N2 (Hadith). We find that the simplest constraints, definite noun phrase & exact match constraints, are the most effective in our case study assuming the gold mentions. We also suggest an approach to compute back pointers in coreference resolution globally.

## Acknowledgment

This work acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number 295703 (FET project "Machine Understanding for interactive StorytElling" (MUSE) `http://www.muse-project.eu/`).

## References

Regina Barzilay and Mirella Lapata. 2006. Aggregation via set partitioning for natural language generation. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL '06, pages 359–366, Stroudsburg, PA, USA. Association for Computational Linguistics.

David I. Beaver. 2004. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3–56.

Eric Bengtson and Dan Roth. 2008. Understanding the value of features for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 294–303, Stroudsburg, PA, USA. Association for Computational Linguistics.

Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL '09, pages 43–48, Stroudsburg, PA, USA. ACL.

Hal Daumé, III, Abhishek Kumar, and Avishek Saha. 2010. Frustratingly easy semi-supervised domain

[11] with gold mentions

adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, DANLP 2010, pages 53–59, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pascal Denis and Jason Baldridge. 2007. Joint determination of anaphoricity and coreference resolution using integer programming. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 236–243, Rochester, New York, April. Association for Computational Linguistics.

Greg Durrett and Dan Klein. 2013. Easy victories and uphill battles in coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, October. Association for Computational Linguistics.

Jenny Rose Finkel and Christopher D. Manning. 2008. Enforcing transitivity in coreference resolution. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Short Papers*, pages 45–48.

M.A. Finlayson and R Hervs. 2010. Ucm/mit indications, referring expressions, and co-reference corpus v1.1 (umirec corpus). MIT CSAIL Work Product.

Mark A. Finlayson, Jeffry R. Halverson, and Steven R. Corman. 2014. The n2 corpus: A semantically annotated collection of islamist extremist stories. The 9th Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–226.

Aria Haghighi and Dan Klein. 2009. Simple coreference resolution with rich syntactic and semantic features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1152–1161, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aria Haghighi and Dan Klein. 2010. Coreference resolution in a modular, entity-centered model. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 385–393, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hudson-D'Zmura. 1988. The structure of discourse and anaphore resolution: The discourse center and the roles of nouns and pronouns. Unpublished doctoral dissertation.

Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, CONLL Shared Task '11, pages 28–34, Stroudsburg, PA, USA. Association for Computational Linguistics.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Alan Prince and Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Wiley-Blackwell.

Manfred Stede. 2011. *Discourse processing*. Morgan & Claypool Publishers.

Veselin Stoyanov, Claire Cardie, Nathan Gilbert, Ellen Riloff, David Buttler, and David Hysom. 2010. Coreference resolution with reconcile. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 156–161, Uppsala, Sweden, July. Association for Computational Linguistics.

Michael J. Toolan. 2009. *Narrative Progression in the Short Story: A Corpus Stylistic Approach*. John Benjamins Publishing.