A corpus study of clause combination

Olga Nikitina
Institut für Computerlinguistik
Universität Heidelberg
nikitina@cl.uni-heidelberg.de

Sebastian Padó
Institut für Computerlinguistik
Universität Heidelberg
pado@cl.uni-heidelberg.de

Abstract

We present a corpus-based investigation of cases of clause combination that can be expressed both through coordination or with subordination. We analyse the data with a two-step computational model which first distinguishes subordination from coordination and then determines the direction for cases of subordination. We find that a wide range of features help with the prediction, notably frequency of predicate participants, presence of adjuncts and sharing of participants between the clause predicates.

1 Introduction

Subordination and coordination are the two primary ways to combine syntactic phrases into sentences. Coordination is a paratactic way of combining constituents (typically) of the same category, where the whole construction has the same type as its daughters. Coordination stands in opposition to subordination, where one constituent is syntactically dependent on another, and where the whole construction has the same type as only one of its daughters, the head daughter. Figure 1 shows examples of both constructions.

In subordinate structures, the dependent constituent can occupy the position of either argument or adjunct. In this paper, we ignore cases of subordinated argument clauses, since their occurrence is mandated mainly by the subcategorization properties of the main clause predicate. Instead, we focus on subordinated clauses that are adjuncts of the main clauses, such as gerund constructions or clauses introduced by subordinating conjunctions (*when, because, …*). Such adjunctive clauses typically describe independent events that stand in some relation to the main clause event.

For such cases, the question arises what determines the speaker's choice between subordination and coordination. It is discussed controversially in the literature. Matthiessen and Thompson (1988) argue that subordination and coordination constructions are grammaticalized discourse relations, with coordination representing the paratactic discourse relations such as *Sequence* and adjunctive subordination constructions representing subjective hypotactic discourse relations such as *Condition* or *Circumstance*. Goldsmith (1985) and Culicover and Jackendoff (1997) list instances of coordination constructions with semantics that is different from that of a sequence, and demonstrate that coordination constructions can express, for example, condition (cf. the example in Figure. 1, a simplified version of their original example *You drink one more can of beer and I'm leaving*). Similarly, there are coordination constructions with causal, concessive, and other meanings.

The goal of our study is to analyze a broader range of factors and their influence on the coordination/subordination choice. To this end, we perform a corpus-based analysis that investigates properties of predicates (and the events which they express) that correlate with the choice between subordination and coordination.

Our study considers three groups of features that are useful for the prediction of clause combination type between two clauses: the frequency and recency of predicates' participants, the presence of adjuncts and the sharing of semantic arguments between predicates. We show that the subordination-coordination choice is not based exclusively on discourse factors, but also correlates with the presence of common participants of predicates as well as with the number and type of predicate modifiers.

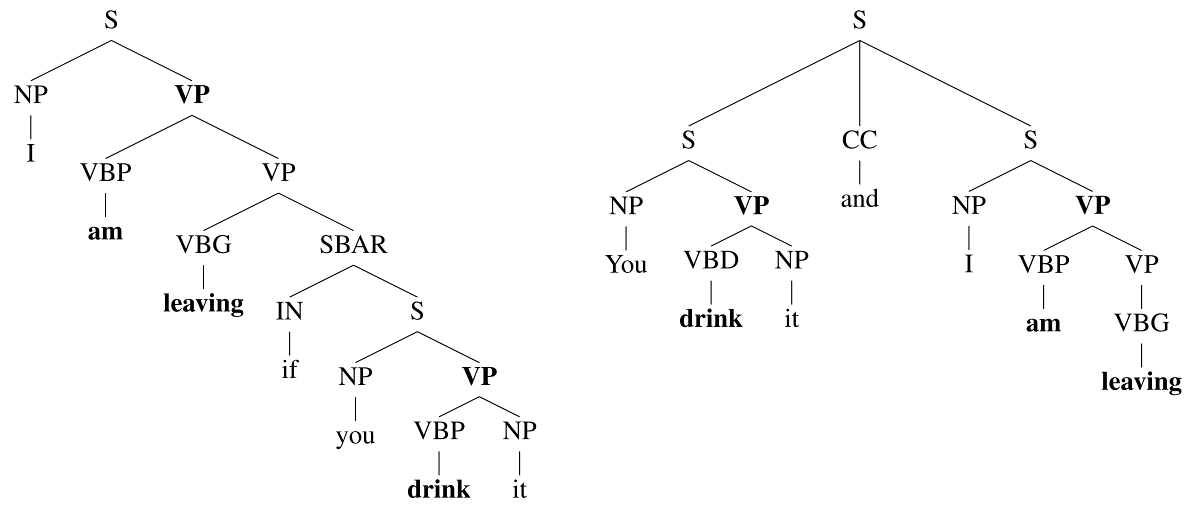

Figure 1: Examples of subordination (left) and coordination (right). The VP projections of the predicates are marked in boldface.

Plan of the paper. In Section 2 we formulate the task and present our method of feature evaluation. In Section 3, we present the features and analyze their usefulness for the prediction tasks. Section 4 contains the evaluation of our model against a majority-case baseline and a model that relies on morphological features. We relate our research to previous studies in Section 5 and summarize the results and give conclusions in Section 6.

2 Method

We adopt a corpus-based method to study the coordination vs. subordination choice. We use the OntoNotes corpus (Pradhan et al., 2007) to extract cases of coordination and subordination and analyze various classes of features that can be suspected in correlating with the coordination/subordination distinction. We evaluate the predictions of our classifiers against the relations between clauses in the original text, which we treat as the gold standard.

This section formulates our task more precisely. We begin by describing our operationalization of the terms "coordination" and "subordination". Then we propose a way to estimate the correlations between the type of clause combination and features of the clauses. Finally, we describe the corpus that we exploit in our experiments.

2.1 Operationalizing Subordination and Coordination

For the analysis, we need to define subordination and coordination constructions in terms of Penn Treebank parse trees and other layers of corpus annotation. For all predicates marked in the PropBank layer of the corpus annotation we define their VP projections (see examples in Figure 1). If the VPs (or their dominating S-nodes) are located at the same level in the tree and if their mother node is also a VP (or an S, respectively), and if they are linked by a coordination conjunction (a word with the part-of-speech CC), we consider these pairs of predicates as *coordination constructions*. *Subordination constructions* are pairs of predicates where one VP is embedded in another VP. More specifically, we define X to be subordinate to Y if there are exactly two VPs on the path between X and Y which correspond to the projections of X and Y, respectively.[1]. Additionally, we restrict our attention to subordinate clauses that are adjucts as motivated in Section 1. We use the PropBank annotation layer to filter out all pairs where one of the

[1]For a deeper analysis of syntactic and semantic differences between coordinate and subordinate structures see, for example, Haspelmath (2004).

predicates is marked as an argument of the second predicate, or where a predicate is the main verb of a sentence that occupies a position of an argument. The semantic information from PropBank allows us to distinguish between syntactically identical argument and adjunct clauses, e.g. “*I liked singing*” and “*I stood singing*”. We also exclude all relative clauses and other clausal noun modifiers. For this, we make sure that the path between predicate projections does not contain NPs.

2.2 Statistical Model and Evaluation

Given a pair of predicates (p_1,p_2), we are faced with two binary decision tasks: (a) p_1 and p_2 can be coordinated or one of them can be subordinated to another, and (b) if the predicates form a subordination construction, either of them can be the main predicate. The first task is the task of prediction the clause combination type, the second is the task of predicting the direction of subodination.

For the purposes of computational modeling, we treat these two decisions as independent and sequential. For each task, we train a binary classifier on sets of features that can influence the clause combination type. More specifically, we make use of logistic regression models, a method that in the past furnished estimates of the importance of different factors in explaining linguistic variation, see e.g. Bresnan et al. (2007) or Hayes and Wilson (2008).

Formally, logistic regression models assume that datapoints consist of a set of predictors x and a binary response variable y. They have the form

$$p(y=1) = \frac{1}{1+e^{-z}} \text{ with } z = \sum_i \beta_i x_i \tag{1}$$

where p is the probability of a datapoint x, β_i is the weight assigned to the predictor x_i. Model estimation sets the parameters β so that the likelihood of the observed (training) data is maximized.

We construct one classifier for each task and for each response variable: $\text{subord-type}(p_1, p_2)$ computes the probability for p_1 and p_2 being linked by a subordination relation, $\text{coord-type}(p_1, p_2)$ computes the probablity for the two predicates to be coordinated, $\text{subord-dir}(p_1, p_2)$ calculates the probablity the probability for p_2 being subordinated to p_1, and $\text{subord-type}(p_2, p_1)$ computes the probability that p_2 dominates p_1. For the first task, we compute the outcome as $\arg\max\{\text{coord-type}(p_1, p_2), \text{subord-type}(p_1, p_2)\}$, and for the second task as $\arg\max\{\text{subord-dir}(p_1, p_2), \text{subord-dir}(p_2, p_1)\}$, respectively. Note that we assume that coordination is a symmetrical relation and we do neither predict nor utilize the linear order of predicates in the original sentence.

Within this scenario, we perform an analysis of individual features and feature groups according to standard practice in the statistics community by considering the effect of features on the models' residual deviance. Residual deviance describes the ratio of the likelihood of the data under a “saturated” model to the likelihood of the data under the actual model (Baayen, 2011). Large decreases in residual deviance that result from the addition of a feature indicate that the feature has substantially increased the ability of the model to explain the data. The statistical significance of the decrease can be determined with the chi-square test.

Since this analysis considers only the training set, it is amenable to overfitting. We therefore add a second kind of analysis that evaluates the model trained on an unseen test set. As the figure of merit, we use simple accuracy (percentage of correctly predicted clause combination types) and compare it against two different baselines (Section 4).

2.3 Corpus

We run our training and testing on the release 4.0 of the OntoNotes corpus (Pradhan et al. (2007)). It contains several layers of annotation, including the PropBank annotation of predicate-argument structures (Palmer et al., 2005), Penn Treebank-style parses (Marcus et al., 1993), and a coreference annotation layer (BBN Technologies, 2007). The WSJ sections 00, 02-04, 09-12, 14, and 17 are used for training and section 20 is used for testing. There are in total 732 documents in the training part of the corpus and 76 in the testing subcorpus. Documents include an average of 46.4 sentences and 109.3 predicates, respectively.

Training corpus		Testing corpus	
Subordinate pairs	7691	Subordinate pairs	736
Coordinate pairs	2187	Coordinate pairs	182
Other pairs	625	Other pairs	61
Total number of pairs	10530	Total number of pairs	979

Table 1: Training and testing corpora

Our training and testing corpora contain three types of predicate pairs. Pairs of the first type are those that joined by the subordination relation, and pairs of the second type are coordination pairs. Third type pairs are those that resemble coordination, but are not linked by any conjunction. We do not consider these cases. Table 1 shows the most important statistics.

Note that while the labels for our first task (subordination vs. coordination, cf. Section 2.2) are "read off" the corpus instances, the relation between the predicates in each subordination pair is not correlated with the actual order of the predicates in the text. In our representation of the data, we broke down all subordination pairs randomly in two classes of comparable size (3833 and 3858 pairs in each class, respectively). In one case all features for p_1 correspond to the features of the main verb, and in another class all features of p_1 describe the dependent verb.

3 Features and Feature Analysis

Table 2 lists the features that we consider in our study. Most features describe *predicates* p_1 or p_2, i.e., the head verbs of adjacent clauses. Each predicate describes an event, typically with one or more participants. Formally, we model *participants* as collections of coreferent NPs (as manually annotated on the coreference level of the corpus). The relationship between participants and predicates is captured on the level of *semantic roles* as annotated on the predicate-argument (PropBank) level of the corpus (e.g., ARG0 is the agent, ARG1 is the patient). Participants can fill more than one role for one predicate, or roles of more than one predicate. In these cases, we talk about *sharing* of participants.

Our features fall into three groups:

Salience features exploit the idea that the discourse status of events is reflected in their syntactic position in the sentence (Matthiessen and Thompson, 1988): key events that are necessary for the understanding of the story cannot be expressed as subordinate clauses. If this holds, it could be expected that such events have more salient participants of the discourse as arguments, and that their discourse status is at least partially determined by the salience of their participants. We assess the salience of participants with a total of 20 features, using some of the features used in anaphora resolution tasks: participant frequency and distance to the previous mention (see Chiarcos (2011) and Mitkov (1998), among others). Participant frequency should show how salient the participant is for the overall document. The distance to the previous mention helps to trace down smaller topics and characterize the participant's role in the local discourse.

Adjunct features cover the expression of adjuncts of the predicates. This group is designed to test whether presence of non-clausal modifiers of predicates influence their syntactic combination. The idea behind including these features is two-fold: on the one hand, they might account for the size of the clauses that should be combined. On the other hand, they might give a clue to us, what properties of events are referred to in the context of the two clause combinations.

Shared participant features test the hypothesis that clauses are syntactically connected because they share content, namely they describe events with identical participants. It was shown before that mentions of same entities may be employed to detect global discourse structure (see Section 5), therefore, it might be possible that they also act on a more local level.

Feat. id	Feature description
Salience features	
f_{1-2}	Number of mentions of the most frequent participant of a predicate
f_{3-4}	Average frequency of all participants of a predicate
f_{5-6}	Average participant frequency, discounted by log of document length in clauses
f_{7-8}	Number of mentions of the most frequent participant, discounted by log of document length in clauses
f_{9-10}	Average participant frequency, discounted by log of number of participants in the document
f_{11-12}	Number of mentions of the most frequent participant, discounted by log of number of participants in the document
f_{13}	Are the most frequent participants of the two predicates equally frequent?
f_{14}	Is the most frequent participant of p_1 mentioned more often than that of p_2?
f_{15-16}	Have any of the participants of a predicate been mentioned previously in text
f_{17-18}	Distance to the previous mention of the participant, minimum over all participants
f_{19}	Has the most recently mentioned participant of p_2 appeared in the document is the same sentence as the most recent participant of p_1?
f_{20}	Has the most recently mentioned participant of p_2 appeared in the document in a sentence that comes before the sentence, where the most recent participant of p_1 was mentioned for the first time?
Adjunct features	
f_{21-22}	Number of adjuncts (of any type)
f_{23-24}	Number of temporal adjuncts
f_{25-26}	Number of locative adjuncts
f_{27-28}	Number of purpose adjuncts
f_{29-30}	Number of causal adjuncts
f_{31-32}	Number of manner adjuncts
Shared participant features	
f_{33}	Are there any shared participants between the predicates?
f_{34}	Number of shared participants between the predicates
f_{35-36}	Does the agent of a predicate coincide with other participants?
f_{37-38}	Does the patient of a predicate coincide with other participants?
f_{39}	Does the agent of p_1 coincide with the patient of p_2?
f_{40}	Does the patient of p_1 coincide with the agent of p_2?
f_{41}	Do the agents of predicates coincide?
f_{42}	Do the patients of predicates coincide?

Table 2: Features for clause combination type prediction. Features with double feature id (e.g. f_{3-4}) are computed separately for each predicate (one for the predicate p_1, one for the predicate p_2)

In the rest of this section, we model the feature groups individually to assess their contribution overall and in terms of single features (cf. Section 2.2).

3.1 Salience Features

The results for predicting subordination/coordination based on salience features are given in Table 3. We find that of all salience features, only features that estimate participant frequency and novelty are useful for the prediction of clause combination type. In fact, predicates with equally frequent participants are more likely to be coordinated than form a subordination construction. This feature has a far greater impact on the model performance than any other feature.

In subordination constructions, verbs with old participants are dispreferred in subordinated positions, while simultaneously verbs with overall more frequent participants are more likely to be dependent on other verbs. We interpret this result, surprising at first sight, to mean that "early" mentions of frequent participants are often found in subordinate clauses. Indeed, this situation is common for news articles, where main participants of the news story are introduced in the first sentence. In the following example, the NP *the American Bar Association* that will be subsequently mentioned in the text several times is first introduced in the subordinate clause: *The Bush administration's nomination of Clarence Thomas to a seat*

Feature id and description	Response variable					
	Clause comb. type is subord.			*p_2 subordinated to p_1*		
	Coefficient	$-\Delta$ RD	Sig.	Coefficient	$-\Delta$ RD	Sig.
Intercept	1.58196	–	–	0.214595	–	–
f_3: Average participant freq. of p_1	-0.04649	35.288	***	0.008190	0.977	
f_4: Average participant freq. of p_2	-0.03265	6.259	*	0.004631	0.044	
f_{15}: A participant of p_1 is not new	0.06243	20.348	***	0.501011	4.978	*
f_{16}: A participant of p_2 is not new	0.03532	4.398	*	-0.503871	16.493	***
f_{13}: Equally frequent participants	-0.59391	180.477	***	0.268878	0.002	
f_{14}: Participants of p_1 are more frequent	-0.02845	0.064		-0.547673	33.708	***
f_{20}: Participants of p_1 are more recent	0.31972	9.994	**	0.008335	0.028	
f_{19}: Equally recent participants	-0.02007	0.031		0.034358	0.105	.

Table 3: Frequency-related features ($-\Delta$ RD: drop in residual deviance; Sig.: Statistical significance, .: $p<0.1$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

on the federal appeals court here received a blow this week when the American Bar Association gave Mr. Thomas only a "qualified" rating, rather than "well qualified".

In sum, the analysis of salience features shows that the discourse status of participants correlates with the syntactic structure of the sentences only mildly. They may be relevant mostly for the prediction of clause combination, but not for the prediction of direction of subordination.

3.2 Adjunct Features

The model that explores the influence of expressed adjuncts is given in Tables 4. The model includes features that describe the number of expressed non-clausal adjuncts. Verbs with temporal and, in particular, locative modifiers tend to be coordinated. Presence of causal adjuncts, on the other hand, increases the probablity of a subordination relation.

A possible explanation is that texts that involve descriptions of locations of different objects and events include more coordination constructions. When the text discusses cause and effects, it is more likely to contain subordination constructions which allow a more precise expression of the semantic relation through subordinating conjunctions. This idea extends the hypothesis that RST relations such as *Cause* are grammaticalized as subordination relations by suggesting that subordination is also likely to be used for other phrasal adjucts of clauses with causal adjuncts.

Within the category of subordination constructions, main clauses tend to contain less adjuncts than subordinated clauses. Clauses that are "heavy" with adjuncts generally reside lower in the syntactic tree; this may be due to considerations similar to those involved in the "heavy NP shift" within clauses (Ross, 1967). The presence of locative adjuncts is the only feature that has a significant effect on the prediction and that runs counter to this pattern.

3.3 Shared Participant Features

Table 5 shows an analysis of the participant features. The features f_{39-40} are most useful for the prediction of dependency direction in the subordination construction. Sharing of the patient of p_1 with the agent of p_2 (f_{39}) is a very strong indicator that p_1 assumes the position of main verb. On the other hand, coinciding agents (f_{41}) and patients (f_{42}) suggest that the verbs are most likely coordinated. The coefficient for the feature f_{34} suggests that the more participants two predicates have in common, the more likely it is that they form a subordination construction.

However, in our experiments we found out that the feature that indicates the presence of shared participants (f_{33}) has the opposite impact on the prediction of clause combination type. In fact, if we treat the f_{34} as a discrete variable, we obtain clearer results. Exactly one common participant increases the chances that the predicates are coordinated, but as the number of shared participants grows, the subordination becomes a more probable alternative (see Table 6).

Feature id and description	Response variable					
	Subordinative combination			*p_2 subordinated to p_1*		
	Coefficient	$-\Delta$ RD	Sig.	Coefficient	Drop $-\Delta$ RD	Sig.
Intercept	1.06692	–	–	0.007055	–	–
f_{23}: Number of temp. adjuncts of p_1	-0.07231	3.5674	.	-0.246345	20.2885	***
f_{24}: Number of temp. adjuncts of p_2	-0.13936	8.3621	**	0.230160	17.5562	***
f_{25}: Number of loc. adjuncts of p_1	-0.36687	19.9467	***	0.127297	1.5190	
f_{26}: Number of loc. adjuncts of p_2	-0.25236	7.6491	**	-0.201170	3.9508	*
f_{27}: Number of purp. adjuncts of p_1	0.39889	0.9586		0.364775	0.9158	
f_{28}: Number of purp. adjuncts of p_2	0.79290	2.0314		-0.437820	0.9268	
f_{29}: Number of cause adjuncts of p_1	0.84954	11.4742	***	-0.028193	0.0160	
f_{30}: Number of cause adjuncts of p_2	0.32878	2.8584	.	0.410857	5.2560	*
f_{21}: Number of manner adjuncts of p_1	0.03634	0.0960		-0.190564	5.8010	*
f_{32}: Number of manner adjuncts of p_2	-0.18875	5.8788	*	0.040456	0.2218	

Table 4: Features characterizing non-clausal adjuncts of the predicates ($-\Delta$ RD: drop in residual deviance; Sig.: Statistical significance, .: $p<0.1$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Feature id and description	Response variable					
	Subordinative combination			*p_2 subordinated to p_1*		
	Coefficient	$-\Delta$ RD	Sig.	Coefficient	$-\Delta$ RD	Sig.
Intercept	1.31168	–	–	-0.0003424	–	–
f_{34}: Number of shared participants	0.18660	205.41	***	0.0137141	0.326	
f_{39}: The agent of p_1 is the patient of p_2	-0.59652	8.02	**	-1.5567667	113.404	***
f_{40}: The agent of p_2 is the patient of p_1	-0.68431	2.23		1.6358964	100.761	***
f_{41}: The agents are the same entity	-1.72330	391.99	***	-0.1119343	1.651	
f_{42}: The patients are the same entity	-1.66920	153.62	***	0.1537697	0.652	

Table 5: Features describing the sharing of participants ($-\Delta$ RD: drop in residual deviance; Sig.: Statistical significance, .: $p<0.1$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Thus, there is a non-linear dependency between participant sharing features and clause combination type. Subordination and coordination constructions have distinct patterns of participant sharing. For coordination, it is often exactly one participant that occupies the same semantic role in the frames of both predicates. More shared participants mean that the verbs are more likely to be subordinated. These constructions are distinguished by the tendency to share participants between the patient of the main verb and the agent of the dependent verb.

We have also tested models with less specific features f_{35-38} and noticed that in subordination constructions, the dependent verb is generally likely to share its agent with one of other participants of the main predicate. In the task of the prediction of clause combination type, having shared participants in almost any role is more likely for coordination constructions, with higher coefficients for agent sharing (f_{35-36}).

4 Prediction of Clause Combination Type

In this Section, we build a model that incorporates all the features that we have discussed in the previous section and use it to predict the test portion of our dataset.

This model (the *Semantic/Discourse Model*), is created on the basis of the most successful features according to our previous analyses. Specifically, it includes 14 features which (a) estimate and compare the frequency of the participants of the verbs in the pair ($f_{3,13,14}$), which (b) register whether any of the participants were mentioned previously ($f_{15,16,20}$), (c) the number of expressed temporal, locative and causal adjuncts ($f_{23-26,29}$), and (d) that report on whether participants are shared between agent and patient roles of the two predicates ($f_{34,39-42}$). We build one classifier for each class and combine them as described in Section 2.2.

Feature id and description	Response variable					
	Subordinative combination			*p_2 subordinated to p_1*		
	Coefficient	$-\Delta$ RD	Sig.	Coefficient	$-\Delta$ RD	Sig.
Intercept	1.45711	–	–	0.004735	–	–
f_{34}=1: One shared participants	-0.99505			-0.013430		
f_{34}=2: Two shared participants	0.05641			-0.261963		
f_{34}=3: Three shared participants	0.92524	766.36	***	0.090377	9.938	.
f_{34}=4: Four shared participants	0.72261			0.004918		
f_{34}=5: Five shared participants	-0.50095			0.682436		
f_{39}: The agent of p_1 is the patient of p_2	-0.50048	4.90	*	-1.493254	112.868	***
f_{40}: The agent of p_2 is the patient of p_1	-0.61984	0.37		1.710374	101.606	***
f_{41}: The agents are the same entity	-1.50994	265.26	***	0.006720	0.100	
f_{42}: The patients are the same entity	-1.34141	94.45	***	0.233702	1.455	

Table 6: Features describing the sharing of participants ($-\Delta$ RD: drop in residual deviance; Sig.: Statistical significance, .: $p<0.1$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Feature id and description	Response variable					
	Subordinative combination			*p_2 subordinated to p_1*		
	Coefficient	$-\Delta$ RD	Sig.	Coefficient	$-\Delta$ RD	Sig.
Intercept	0.94877	–	–	0.002226	–	–
p_1 is a gerund	0.36027			-1.702426		
p_1 is an infinitive	0.11551	51.514	***	-1.187537	514.90	***
p_1 is a participle	-0.53580			-0.979405		
p_2 is a gerund	0.26939			1.835103		
p_2 is an infinitive	0.11392	24.017	***	1.014216	530.34	***
p_2 is a participle	-0.45433			1.154144		

Table 7: Features for the Morphological Model: verb form of p_1 and verb form of p_2. Each value is assigned a coefficient, but the drop in residual deviance is computed at the feature level. ($-\Delta$ RD: drop in residual deviance; Sig.: Statistical significance, .: $p<0.1$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

We compare our Semantic/Discourse model to two other models. The first one, *Majority Baseline*, assigns every pair to the most frequent class. For the first task (subordination vs. coordination), this is subordination (75% of instances); for the second, this is subordination of p_1 under p_2 (52% of instances).

Our second point of comparison is the *Morphological Model*. This model is based on just two features, namely the morphological forms of the two verbs. These features allow the model to solve the second task in cases when subordinated predicate has a non-finite form. However, from our point of view, this model is not fit for our purposes since it uses information which from a generation perspective is not yet available at the point in time when syntactic decisions have to be made.

Table 7 lists these features in a similar manner to the semantic and discourse features used in Section 3. Both features are modelled as factors with four levels each (the three listed ones plus the base level of finite verb).[2]

The results of applying these models to our OntoNotes test set are shown in Table 8. For the first task, the accuracy of the Majority Baseline classifier corresponds to the proportion of the majority classes in the dataset (0.75). The Morphological Model follows the Baseline in assigning all cases to the subordination case and thus achieves the same overall accuracy. On the second task, it improves substantially over the baseline (accuracy 0.606) due to correct predictions in cases where subordinated clauses have non-finite predicates. At the same time, when both predicates have finite form (which is the majority of our data) the classifier cannot make any informed decision. However, although its intercept feature is very close to zero, it still has a little bias towards one of the classes, which is mirrored in the accuracy of prediction on different subsets of the data (0.921 vs. 0.316).

Our Semantics/Discourse Model is able to improve over the two other models for the subordination

[2]Consequently, there are three coefficients but just one drop in residual deviance resulting from the addition of the feature.

Model	Subordination vs. coordination			Direction of subordination		
	Overall	Subord.	Coord.	Overall	p_2 subord. to p_1	p_1 subord. to p_2
Majority Baseline	0.752	1.000	0.000	0.520	0.000	1.000
Morphological Model	0.752	1.000	0.000	0.606	0.921	0.316
Semantics/Discourse Model	0.779	0.946	0.368	0.576	0.668	0.420

Table 8: Prediction accuracy for the two tasks (Task 1: subordination vs. coordination; Task 2: direction of subordination) in terms of overall accuracy and class-specific accuracy.

vs. coordination task by learning how to recognize at least some cases of coordination. Concerning the direction of subordination, it improves over the baseline (0.576). While it does not achieve the overall accuracy of the Morphological Model, it is more balanced over the two classes. Also, recall that the good performance of the Morphological Model is due to its use of verb form information which is arguably unavailable at the decision time.

5 Related Work

The choice between subordination and coordination is related to work on various aspects of discourse and beyond in computational linguistics.

Rhetorical Relations. The closest area to our work consists of investigations of discourse relations in the context of Rhetorical Structure Theory (Mann and Thompson, 1988). Most studies in this area are primarily concerned with appropriate choice and positioning of the discourse cue, barely considering the differences between syntactic status of clauses to be combined. However, (Taboada, 2006) shows that some rhetorical relations are often expressed without any discourse cue, and such parameters of sentence structure as the order of phrases and their syntactic mode of combination become significant for the expression of rhetoric relation. There are several studies that consider syntactic means of expression of particular rhetorical relations. In particular, Grote et al. (1997) describe how syntactic structure and ordering of clauses correspond to the pragmatic subtypes of the *Concession* relation. Pitler et al. (2009) show that pairs of words taken from sentences linked by discourse relations, as well as Levin classes of verbs of the sentences and sentiment polarity information is useful for the prediction of implicit relations. The same authors also look into various entity-based features and show that again lexical information about mentioned entities correlates with the choice of discourse relation. In contrast, we focus on the correlation between the syntactic structure and the properties of events directly, since both types of clause combination may be used to encode the same rhetoric relation. We think that while the influence of pragmatic factors investigated by Grote et al. (1997) may be significant, we chose to explore other types of features in this study.

Lexical Models of Coherence Another direction of research of coherence relations within discourse is represented by Barzilay and Lapata (2008). They show that coherent discourse is characterized by chains of mentions of same entities. Hearst (1997) show that event chains that are formed only by the mentions of the same lexical item mirror the global structure of texts and can be used for discourse segmentation. The “shared participant” features that we use are similar to the approach of these studies. However, our work shows that coordination and subordination form distinct patterns of entity mentions which can be used to predict local text structure.

Generation and Summarization. We believe that our results may be useful to the natural language generation and summarization communities. In generation, many systems assume overgenerate-and-rank approach to sentence planning (for example, see Stent et al. (2004)). The description of features given in our work may help to create better ranking systems or even direct the generation of complex and compound sentences, in the spirit of Stent and Molina (2009). In summarization, Barzilay and McKeown (2005) present a sentence fusion technique for multidocument summarization which needs to restructure sentences to improve text coherence. Restructuring is currently done without regard to the underlying discourse structure. We believe that the features that we have identified can introduce a bias towards more appropriate structures during sentence fusion.

6 Conclusions

In this paper, we have reported on an examination of various semantic and discourse structure-based factors and their effect on the choice of clause combination (subordination vs. coordination) and the direction of relation within subordination pairs. On a dataset of clause pairs extracted from the OntoNotes corpus, our analysis led to the following results:

- The salience of events and their participants is connected with the syntactic position of corresponding clauses in the tree. However, in order to occupy the dominating position in the syntactic structure, the event only has to be more prominent than another event with which it forms a pair. It does not need to be the key, mainline event of the story.
- The presence of adjuncts of different types has an effect on the clause combination preferences. Locative adjuncts are different from other types of adjuncts and clauses in that they seem to support coordination more than subordination. On the other hand, the presence of causal adjuncts increases the likelihood of subordination constructions.
- Participant sharing between different argument positions of predicate is one of the decisive factors in the prediction of clause combination type. Coordination constructions are more likely to share one participant between same semantic roles of the predicates, whereas the in case of subordination participants are shared between patient and agent positions.

In sum, we find that the choice between subordination and coordination is not determined by "global" discourse factors alone, but also by the lexical and structural properties of the participating predicates and their immediate context. Moreover, the two prediction tasks involve different, often complimentary features. We interpret this as evidence for a richer, more interactive account of clause structuring in discourse context than previous work has suggested.

References

Baayen, H. (2011). *Analyzing Linguistic Data*. Cambridge University Press.

Barzilay, R. and M. Lapata (2008). Modeling Local Coherence: An Entity-Based Approach. *Computational Linguistics 34*(1), 1–34.

Barzilay, R. and K. McKeown (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics 31*(3), 297–328.

BBN Technologies (2004-2007). *Co-reference guidelines for English OntoNotes*. BBN Technologies.

Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 69–94. Royal Netherlands Academy of Science.

Chiarcos, C. (2011). Evaluating salience metrics for the context-adequate realization of discourse referents. In *Proceedings of the 13th European Workshop on Natural Language Generation*, Nancy, France, pp. 32–43.

Culicover, P. W. and R. Jackendoff (1997). Semantic subordination despite syntactic coordination. *Linguistic Inquiry 28*(2), 195–217.

Goldsmith, J. (1985). A principled exception to the coordinate structure constraint. In *Proceedings of the 21st Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Grote, B., N. Lenke, and M. Stede (1997). Ma(r)king concessions in English and German. *Discourse Processes 24*, 87–117.

Haspelmath, M. (Ed.) (2004). *Coordinating constructions*. Amsterdam: Benjamins.

Hayes, B. and C. Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry 39*, 379–440.

Hearst, M. A. (1997). Texttiling: segmenting text into multi-paragraph subtopic passages. *Computational Linguistics 23*(1), 33–64.

Mann, W. C. and S. A. Thompson (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text 8*(3), 243–281.

Marcus, M., B. Santorini, and M. A. Marcinkiewicz (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics 19*(2), 313–330.

Matthiessen, C. and S. A. Thompson (1988). The structure of discourse and 'subordination'. In J. Haiman and S. A. Thompson (Eds.), *Clause combining in grammar and discourse*. Amsterdam: Benjamins.

Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *Proceedings of ACL/COLING*, Montreal, Canada, pp. 869–875.

Palmer, M., D. Gildea, and P. Kingsbury (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics 31*(2), 71–106.

Pitler, E., A. Louis, and A. Nenkova (2009). Automatic sense prediction for implicit discourse relations in text. In *ACL/AFNLP*, pp. 683–691.

Pradhan, S., E. H. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel (2007). OntoNotes: A Unified Relational Semantic Representation. In *Proceedings of the First IEEE International Conference on Semantic Computing*, pp. 517–526.

Ross, J. (1967). *Constraints on variables in syntax*. Ph. D. thesis, MIT.

Stent, A. and M. Molina (2009). Evaluating automatic extraction of rules for sentence plan construction. In *Proceedings of the SIGDIAL 2009 Conference*, London, UK, pp. 290–297.

Stent, A., R. Prasad, and M. Walker (2004). Trainable sentence planning for complex information presentation in spoken dialog systems. In *In Proc. of the Annual Meeting of the Association for Computational Linguistics*, pp. 79–86.

Taboada, M. T. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics. 38*(4), 567–592.