Incorporating Information Status into Generation Ranking

Aoife Cahill and Arndt Riester
Institut für Maschinelle Sprachverarbeitung (IMS)
University of Stuttgart
70174 Stuttgart, Germany
`{aoife.cahill,arndt.riester}@ims.uni-stuttgart.de`

Abstract

We investigate the influence of information status (IS) on constituent order in German, and integrate our findings into a log-linear surface realisation ranking model. We show that the distribution of pairs of IS categories is strongly asymmetric. Moreover, each category is correlated with morphosyntactic features, which can be automatically detected. We build a log-linear model that incorporates these asymmetries for ranking German string realisations from input LFG F-structures. We show that it achieves a statistically significantly higher BLEU score than the baseline system without these features.

1 Introduction

There are many factors that influence word order, e.g. humanness, definiteness, linear order of grammatical functions, givenness, focus, constituent weight. In some cases, it can be relatively straightforward to automatically detect these features (i.e. in the case of definiteness, this is a syntactic property). The more complex the feature, the more difficult it is to automatically detect. It is common knowledge that information status[1] (henceforth, IS) has a strong influence on syntax and word order; for instance, in inversions, where the subject follows some preposed element, Birner (1994) reports that the preposed element must not be newer in the discourse than the subject. We would like to be able to use information related to IS in the automatic generation of German text. Ideally, we would automatically annotate text with IS labels and learn from this data. Unfortunately, however, to date, there has been little success in automatically annotating text with IS.

---

[1] We take *information status* to be a subarea of *information structure*; the one dealing with varieties of givenness but not with contrast and focus in the strictest sense.

We believe, however, that despite this shortcoming, we can still take advantage of some of the insights gained from looking at the influence of IS on word order. Specifically, we look at the problem from a more general perspective by computing an asymmetry ratio for each pair of IS categories. Results show that there are a large number of pairs exhibiting clear ordering preferences when co-occurring in the same clause. The question then becomes, without being able to automatically detect these IS category pairs, can we, nevertheless, take advantage of these strong asymmetric patterns in generation. We investigate the (automatically detectable) morphosyntactic characteristics of each asymmetric IS pair and integrate these syntactic asymmetric properties into the generation process.

The paper is structured as follows: Section 2 outlines the underlying realisation ranking system for our experiments. Section 3 introduces information status and Section 4 describes how we extract and measure asymmetries in information status. In Section 5, we examine the syntactic characteristics of the IS asymmetries. Section 6 outlines realisation ranking experiments to test the integration of IS into the system. We discuss our findings in Section 7 and finally we conclude in Section 8.

2 Generation Ranking

The task we are considering is generation ranking. In generation (or more specifically, surface realisation) ranking, we take an abstract representation of a sentence (for example, as produced by a machine translation or automatic summarisation system), produce a number of alternative string realisations corresponding to that input and use some model to choose the most likely string. We take the model outlined in Cahill et al. (2007), a log-linear model based on the Lexical Functional Grammar (LFG) Framework (Kaplan and Bresnan, 1982). LFG has two main levels of represen-

CS 1:   ROOT:1458

              CProot[std]:1451   PERIOD:397

        DP[std]:906   Cbar:1448   .:389

     DPx[std]:903   Cbar-flat:1436

  D[std]:593 NP:738   V[v,fin]:976   PP[std]:2081

   die:34 N[comm]:693 Vx[v,fin]:973   PPx[std]:2072

      Behörden:85 warnten:117 P[pre]:1013 DP[std]:1894

                              vor:154   DPx[std]:1956

                                        NP:1952

                           AP[std,+infl]:1946 N[comm]:1252

                        APx[std,+infl]:1928 Nachbeben:263

                          A[+infl]:1039

                          möglichen:185

"Die Behörden warnten vor möglichen Nachbeben."

```
      PRED    'warnen<[34:Behörde], [263:Nachbeben]>'
              PRED  'Behörde'
      SUBJ    SPEC [DET [PRED 'die']]
           34 CASE nom, NUM pl, PERS 3
              PRED   'vor<[263:Nachbeben]>'
                     PRED    'Nachbeben'
      OBL                             PRED   'möglich<[263:Nachbeben]>'
              OBJ    ADJUNCT          SUBJ   [263:Nachbeben]
                                  185 ATYPE  attributive
                 154        263 CASE dat, NUM pl, PERS 3
      TNS-ASP [MOOD indicative, TENSE past]
  117 TOPIC   [34:Behörde]
```
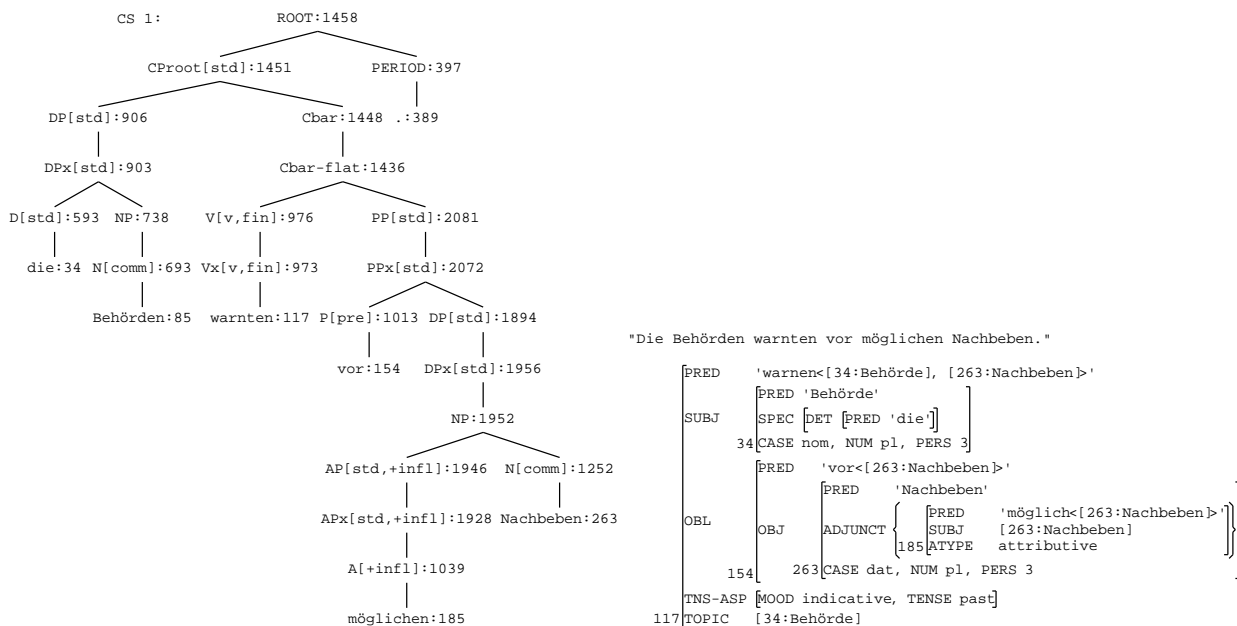
Figure 1: An example C(onstituent) and F(unctional) Structure pair for (1)

tation, C(onstituent)-Structure and F(unctional)-Structure. C-Structure is a context-free tree representation that captures characteristics of the surface string while F-Structure is an abstract representation of the basic predicate-argument structure of the string. An example C- and F-Structure pair for the sentence in (1) is given in Figure 1.

(1) Die Behörden warnten vor möglichen Nachbeben.
     the authorities warned of possible   aftershocks

     'The authorities warned of possible aftershocks.'

The input to the generation system is an F-Structure. A hand-crafted, bi-directional LFG of German (Rohrer and Forst, 2006) is used to generate all possible strings (licensed by the grammar) for this input. As the grammar is hand-crafted, it is designed only to parse (and therefore) generate grammatical strings.[2] The task of the realisation ranking system is then to choose the most likely string. Cahill et al. (2007) describe a log-linear model that uses linguistically motivated features and improves over a simple tri-gram language model baseline. We take this log-linear model as our starting point.[3]

An error analysis of the output of that system revealed that sometimes "unnatural" outputs were being selected as most probable, and that often information structural effects were the cause of subtle differences in possible alternatives. For instance, Example (3) appeared in the original TIGER corpus with the 2 preceding sentences (2).

(2) Denn ausdrücklich ist darin der rechtliche Maßstab der Vorinstanz, des Sächsischen Oberverwaltungsgerichtes, bestätigt worden. Und der besagt: Die Beteiligung am politischen Strafrecht der DDR, der Mangel an kritischer Auseinandersetzung mit totalitären Überzeugungen rechtfertigen den Ausschluss von der Dritten Gewalt.

     'Because, the legal benchmark has explicitly been confirmed by the lower instance, the Saxonian Higher Administrative Court. And it indicates: the participation in the political criminal law of the GDR as well as deficits regarding the critical debate on totalitarian convictions justify an expulsion from the judiciary.'

(3) Man hat aus   der Vergangenheitsaufarbeitung
     one has out of the coming to terms with the past
     gelernt.
     learnt

     'People have learnt from dealing with the past mistakes.'

The five alternatives output by the grammar are:

a. Man hat aus der Vergangenheitsaufarbeitung gelernt.

b. Aus der Vergangenheitsaufarbeitung hat man gelernt.

c. Aus der Vergangenheitsaufarbeitung gelernt hat man.

d. Gelernt hat man aus der Vergangenheitsaufarbeitung.

e. Gelernt hat aus der Vergangenheitsaufarbeitung man.

The string chosen as most likely by the system of Cahill et al. (2007) is Alternative (b). No matter whether the context in (2) is available or the sentence is presented without any context, there seems to be a preference by native speakers for the original string (a). Alternative (e) is extremely marked[4] to the point of being ungrammatical. Alternative (c) is also very marked and so is Alternative (d), although less so than (c) and (e). Alternative (b) is a little more marked than the original string, but it is easier to imagine a preceding context where this sentence would be perfectly appropriate. Such a context would be, e.g. (4).

(4) Vergangenheitsaufarbeitung und Abwiegeln sind zwei sehr unterschiedliche Arten, mit dem Geschehenen umzugehen.

'Dealing with the mistakes or playing them down are two very different ways to handle the past.'

If we limit ourselves to single sentences, the task for the model is then to choose the string that is closest to the "default" expected word order (i.e. appropriate in the most number of contexts). In this work, we concentrate on integrating insights from work on information status into the realisation ranking process.

3 Information Status

The concept of *information status* (Prince, 1981; Prince, 1992) involves classifying NP/PP/DP expressions in texts according to various ways of their being *given* or *new*. It replaces and specifies more clearly the often vaguely used term *givenness*. The process of labelling a corpus for IS can be seen as a means of discourse analysis. Different classification systems have been proposed in the literature; see Riester (2008a) for a comparison of several IS labelling schemes and Riester (2008b) for a new proposal based on criteria from presupposition theory. In the work described here, we use the scheme of Riester (2008b). His main theoretic assumption is that IS categories (for definites) should group expressions according to the contextual resources in which their presuppositions find an antecedent. For definites, the set of main category labels found in Table 1 is assumed.

The idea of resolution contexts derives from the concept of a presupposition trigger (e.g. a definite description) as potentially establishing an

Context resource	IS label
discourse context	D-GIVEN
encyclopedic/ knowledge context	ACCESSIBLE-GENERAL
environment/ situative context	SITUATIVE
bridging context (scenario)	BRIDGING
accommodation (no context)	ACCESSIBLE-DESCRIPTION

Table 1: IS classification for definites

anaphoric relation (van der Sandt, 1992) to an entity being available by some means or other. But there are some expressions whose referent cannot be *identified* and needs to be *accommodated*, compare (5).

(5) [die monatelange Führungskrise der Hamburger Sozialdemokraten]$_{\text{ACC-DESC}}$

'the leadership crisis lasting for months among the Hamburg Social Democrats'

Examples like this one have been mentioned early on in the literature (e.g. Hawkins (1978), Clark and Marshall (1981)). Nevertheless, labelling schemes so far have neglected this issue, which is explicitly incorporated in the system of Riester (2008b).

The status of an expression is ACCESSIBLE-GENERAL (or *unused*, following Prince (1981)) if it is not present in the previous discourse but refers to an entity that is known to the intended recipient. There is a further differentiation of the ACCESSIBLE-GENERAL class into generic (TYPE) and non-generic (TOKEN) items.

An expression is D-GIVEN (or *textually evoked*) if and only if an antecedent is available in the discourse context. D-GIVEN entities are subdivided according to whether they are repetitions of their antecedent, short forms thereof, pronouns or whether they use new linguistic material to add information about an already existing discourse referent (label: EPITHET). Examples representing a co-reference chain are shown in (6).

(6) [Angela Merkel]$_{\text{ACC-GEN}}$ *(first mention)* ... [Angela Merkel]$_{\text{D-GIV-REPEATED}}$ *(second mention)* ... [Merkel]$_{\text{D-GIV-SHORT}}$ ... [she]$_{\text{D-GIV-PRONOUN}}$ ... [herself]$_{\text{D-GIV-REFLEXIVE}}$ ... [the Hamburg-born politician]$_{\text{D-GIV-EPITHET}}$

Indexicals (referring to entities in the environment context) are labeled as SITUATIVE. Definite

---

[4]By marked, we mean that there are relatively few or specialised contexts in which this sentence is acceptable.

items that can be identified within a scenario context evoked by a non-coreferential item receive the label BRIDGING; compare Example (7).

(7) In <u>Sri Lanka</u> haben tamilische Rebellen
     in Sri Lanka have Tamil   rebels
     erstmals   einen Luftangriff [gegen die
     for the first time an airstrike   against the
     Streitkräfte]_BRIDG geflogen.
     armed forces   flown.

     'In Sri Lanka, Tamil rebels have, for the first time, carried out an airstrike against the armed forces.'

In the indefinite domain, a simple classification along the lines of Table 2 is proposed.

Type	IS label
unrelated to context	NEW
part-whole relation to previous entity	PARTITIVE
other (unspecified) relation to context	INDEF-REL

Table 2: IS classification for indefinites

There are a few more subdivisions. Table 3, for instance, contains the labels BRIDGING-CONTAINED and PARTITIVE-CONTAINED, going back to Prince's (1981:236) "containing inferrables". The entire IS label inventory used in this study comprises 19 (sub)classes in total.

4 Asymmetries in IS

In order to find out whether IS categories are unevenly distributed within German sentences we examine a corpus of German radio news bulletins that has been manually annotated for IS (496 annotated sentences in total) using the scheme of Riester (2008b).[5]

For each pair of IS labels X and Y we count how often they co-occur in the corpus within a single clause. In doing so, we distinguish the numbers for "X preceding Y" (= A) and "Y preceding X" (= B). The larger group is referred to as the *dominant order*. Subsequently, we compute a ratio indicating the degree of asymmetry between the two orders. If, for instance, the dominant pattern occurs 20 times (A) and the reverse pattern only 5 times (B), the asymmetry ratio B/A is 0.25.[6]

Dominant order ($\gg$: *"before"*)	B/A	Total
D-GIV-PRO$\gg$INDEF-REL	0	19
D-GIV-PRO$\gg$D-GIV-CAT	0.1	11
D-GIV-REL$\gg$NEW	0.11	31
D-GIV-PRO$\gg$SIT	0.13	17
ACC-DESC$\gg$INDEF-REL	0.14	24
ACC-DESC$\gg$ACC-GEN-TY	0.19	19
D-GIV-EPI$\gg$INDEF-REL	0.2	12
D-GIV-REP$\gg$NEW	0.21	23
D-GIV-PRO$\gg$ACC-GEN-TY	0.22	11
ACC-GEN-TO$\gg$ACC-GEN-TY	0.24	42
D-GIV-PRO$\gg$ACC-DESC	0.24	46
EXPL$\gg$NEW	0.25	30
D-GIV-REL$\gg$D-GIV-EPI	0.25	15
BRIDG-CONT$\gg$PART-CONT	0.25	15
ACC-DESC$\gg$EXPL	0.29	27
D-GIV-PRO$\gg$D-GIV-REP	0.29	18
D-GIV-PRO$\gg$NEW	0.29	88
D-GIV-REL$\gg$ACC-DESC	0.3	26
SIT$\gg$EXPL	0.31	17
D-GIV-PRO$\gg$BRIDG-CONT	0.31	21
D-GIV-PRO$\gg$D-GIV-SHORT	0.32	29
. . .	. . .	
ACC-DESC$\gg$ACC-GEN-TO	0.91	201
SIT$\gg$BRIDG	0.92	23
EXPL$\gg$ACC-DESC	1	12

Table 3: Asymmetric pairs of IS labels

Table 3 gives the top asymmetry pairs down to a ratio of about 1:3 as well as, down at the bottom, the pairs that are most evenly distributed. This means that the top pairs exhibit strong ordering preferences and are, hence, unevenly distributed in German sentences. For instance, the ordering D-GIVEN-PRONOUN before INDEF-REL (top line), shown in Example (8), occurs 19 times in the examined corpus while there is no example in the corpus for the reverse order.[7]

(8) [Sie]_D-GIV-PRO würde auch [bei verringerter
     she   would also at   reduced
     Anzahl]_INDEF-REL jede vernünftige
     number   every sensible
     Verteidigungsplanung sprengen.
     defence planning   blast

     'Even if the numbers were reduced it would blow every sensible defence planning out of proportion.'

5 Syntactic IS Asymmetries

It seems that IS could, in principle, be quite beneficial in the generation ranking task. The problem, of course, is that we do not possess any reliable system of automatically assigning IS labels to unknown text and manual annotations are costly and time-consuming. As a substitute, we identify a list

---

[5]The corpus was labeled by two independent annotators and the results were compared by a third person who took the final decision in case of disagreement. An evaluation as regards inter-coder agreement is currently underway.

[6]Even if some of the sentences we are learning from are marked in terms of word order, the ratios allow us to still learn the predominant order, since the marked order should occur much less frequently and the ratio will remain low.

[7]Note that we are not claiming that the reverse pattern is ungrammatical or impossible, we just observe that it is extremely infrequent.

of morphosyntactic characteristics that the expressions can adopt and investigate how these are correlated to our inventory of IS categories.

For some IS labels there is a direct link between the typical phrases that fall into that IS category, and the syntactic features that describe it. One such example is D-GIVEN-PRONOUN, which always corresponds to a pronoun, or EXPL which always corresponds to expletive items. Such syntactic markers can easily be identified in the LFG F-structures. On the other hand, there are many IS labels for which there is no clear cut syntactic class that describes its typical phrases. Examples include NEW, ACCESSIBLE-GENERAL or ACCESSIBLE-DESCRIPTION.

In order to determine whether we can ascertain a set of syntactic features that are representative of a particular IS label, we design an inventory of syntactic features that are found in all types of IS phrases. The complete inventory is given in Table 5. It is a much easier task to identify these syntactic characteristics than to try and automatically detect IS labels directly, which would require a deep semantic understanding of the text. We automatically mark up the news corpus with these syntactic characteristics, giving us a corpus both annotated for IS and syntactic features.

We can now identify, for each IS label, what the most frequent syntactic characteristics of that label are. Some examples and their frequencies are given in Table 4.

Syntactic feature	Count
D-GIVEN-PRONOUN	
PERS_PRON	39
DA_PRON	25
DEMON_PRON	19
GENERIC_PRON	11
NEW	
SIMPLE_INDEF	113
INDEF_ATTR	53
INDEF_NUM	32
INDEF_PPADJ	26
INDEF_GEN	25
...	

Table 4: Syntactic characteristics of IS labels

Combining the most frequent syntactic characteristics with the asymmetries presented in Table 3 gives us Table 6.[8]

---

[8] For reasons of space, we are only showing the very top of the table.

6 Generation Ranking Experiments

Using the augmented set of IS asymmetries, we design new features to be included into the original model of Cahill et al. (2007). For each IS asymmetry, we extract all precedence patterns of the corresponding syntactic features. For example, from the first asymmetry in Table 6, we extract the following features:

```
PERS_PRON precedes INDEF_ATTR
PERS_PRON precedes SIMPLE_INDEF
DA_PRON precedes INDEF_ATTR
DA_PRON precedes SIMPLE_INDEF
DEMON_PRON precedes INDEF_ATTR
DEMON_PRON precedes SIMPLE_INDEF
GENERIC_PRON precedes INDEF_ATTR
GENERIC_PRON precedes SIMPLE_INDEF
```

We extract these patterns for all of the asymmetric pairs in Table 3 (augmented with syntactic characteristics) that have a ratio >0.4. The patterns we extract need to be checked for inconsistencies because not all of them are valid. By inconsistencies, we mean patterns of the type X precedes X, Y precedes Y, and any pattern where the variant X precedes Y as well as Y precedes X is present. These are all automatically removed from the list of features to give a total of 130 new features for the log-linear ranking model.

We train the log-linear ranking model on 7759 F-structures from the TIGER treebank. We generate strings from each F-structure and take the original treebank string to be the labelled example. All other examples are viewed as unlabelled. We tune the parameters of the log-linear model on a small development set of 63 sentences, and carry out the final evaluation on 261 unseen sentences. The ranking results of the model with the additional IS-inspired features are given in Table 7.

Model	BLEU	Exact Match (%)
Cahill et al. (2007)	0.7366	52.49
New Model (Model 1)	0.7534	54.40

Table 7: Ranking Results for new model with IS-inspired syntactic asymmetry features.

We evaluate the string chosen by the log-linear model against the original treebank string in terms of exact match and BLEU score (Papineni et al.,

Syntactic feature	Type
	Definites
	Definite descriptions
SIMPLE_DEF	simple definite descriptions
POSS_DEF	simple definite descriptions with a possessive determiner (pronoun or possibly genitive name)
DEF_ATTR_ADJ	definite descriptions with adjectival modifier
DEF_GENARG	definite descriptions with a genitive argument
DEF_PPADJ	definite descriptions with a PP adjunct
DEF_RELARG	definite descriptions including a relative clause
DEF_APP	definite descriptions including a title or job description as well as a proper name (e.g. an apposition)
	Names
PROPER	combinations of position/title and proper name (without article)
BARE_PROPER	bare proper names
	Demonstrative descriptions
SIMPLE_DEMON	simple demonstrative descriptions
MOD_DEMON	adjectivally modified demonstrative descriptions
	Pronouns
PERS_PRON	personal pronouns
EXPL_PRON	expletive pronoun
REFL_PRON	reflexive pronoun
DEMON_PRON	demonstrative pronouns (not: determiners)
GENERIC_PRON	generic pronoun *(man – one)*
DA_PRON	"da"-pronouns *(darauf, darüber, dazu, . . .)*
LOC_ADV	location-referring pronouns
TEMP_ADV, YEAR	Dates and times
	Indefinites
SIMPLE_INDEF	simple indefinites
NEG_INDEF	negative indefinites
INDEF_ATTR	indefinites with adjectival modifiers
INDEF_CONTRAST	indefinites with contrastive modifiers *(einige – some, andere – other, weitere – further, . . .)*
INDEF_PPADJ	indefinites with PP adjuncts
INDEF_REL	indefinites with relative clause adjunct
INDEF_GEN	indefinites with genitive adjuncts
INDEF_NUM	measure/number phrases
INDEF_QUANT	quantified indefinites

Table 5: An inventory of interesting syntactic characteristics in IS phrases

Label 1 (+ features)		Label 2 (+ features)		B/A	Total
D-GIVEN-PRONOUN		**INDEF-REL**		0	19
PERS_PRON	39	INDEF_ATTR	23		
DA_PRON	25	SIMPLE_INDEF	17		
DEMON_PRON	19				
GENERIC_PRON	11				
D-GIVEN-PRONOUN		**D-GIVEN-CATAPHOR**		0.1	11
PERS_PRON	39	SIMPLE_DEF	13		
DA_PRON	25	DA_PRON	10		
DEMON_PRON	19				
GENERIC_PRON	11				
D-GIVEN-REFLEXIVE		**NEW**		0.11	31
REFL_PRON	54	SIMPLE_INDEF	113		
		INDEF_ATTR	53		
		INDEF_NUM	32		
		INDEF_PPADJ	26		
		INDEF_GEN	25		
		...			

Table 6: IS asymmetric pairs augmented with syntactic characteristics

2002). We achieve an improvement of 0.0168 BLEU points and 1.91 percentage points in exact match. The improvement in BLEU is statistically significant (p < 0.01) using the paired bootstrap resampling significance test (Koehn, 2004).

Going back to Example (3), the new model chooses a "better" string than the Cahill et al. (2007) model. The new model chooses the original string. While the string chosen by the Cahill et al. (2007) system is also a perfectly valid sentence, our empirical findings from the news corpus were that the default order of generic pronoun before definite NP were more frequent. The system with the new features helped to choose the original string, as it had learnt this asymmetry.

Was it just the syntax?

The results in Table 7 clearly show that the new model is beneficial. However, we want to know how much of the improvement gained is due to the IS asymmetries, and how much the syntactic asymmetries on their own can contribute. To this end, we carry out a further experiment where we calculate syntactic asymmetries based on the automatic markup of the corpus, and ignore the IS labels completely. Again we remove any inconsistent asymmetries and only choose asymmetries with a ratio of higher than 0.4. The top asymmetries are given in Table 8.

Dominant order ($\gg$: *"before"*)	B/A	Total
BAREPROPER$\gg$INDEF_NUM	0	33
DA_PRON$\gg$INDEF_NUM	0	16
DEF_PPADJ$\gg$TEMP_ADV	0	15
SIMPLE_INDEF$\gg$INDEF_QUANT	0	14
PERS_PRON$\gg$INDEF_ATTR	0	12
DEF_PPADJ$\gg$EXPL_PRON	0	12
GENERIC_PRON$\gg$INDEF_ATTR	0	12
REFL_PRON$\gg$YEAR	0	11
INDEF_PPADJ$\gg$INDEF_NUM	0.02	57
DEF_APP$\gg$BAREPROPER	0.03	34
BAREPROPER$\gg$TEMP_ADV	0.04	26
TEMP_ADV$\gg$INDEF_NUM	0.04	25
PROPER$\gg$INDEF_GEN	0.05	20
DEF_GENARG$\gg$INDEF_ATTR	0.06	18
...		...

Table 8: Purely syntactic asymmetries

For each asymmetry, we create a new feature `X precedes Y`. This results in a total of 66 features. Of these 30 overlap with the features used in the above experiment. We do not include the features extracted in the first attempt in this experiment. The same training procedure is carried out and we test on the same heldout test set of 261 sentences. The results are given in Table 9. Finally, we combine the two lists of features and evaluate, these results are also presented in Table 9.

Model	BLEU	Exact Match (%)
Cahill et al. (2007)	0.7366	52.49
Model 1	0.7534	54.40
Synt.-asym.-based Model	0.7419	54.02
Combination	0.7437	53.64

Table 9: Results for ranking model with purely syntactic asymmetry features

They show that although the syntactic asymmetries alone contribute to an improvement over the baseline, the gain is not as large as when the syntactic asymmetries are constrained to correspond to IS label asymmetries (Model 1).[9] Interestingly, the combination of the lists of features does not result in an improvement over Model 1. The difference in BLEU score between the model of Cahill et al. (2007) and the model that only takes syntactic-based asymmetries into account is not statistically significant, while the difference between Model 1 and this model is statistically significant (p < 0.05).

7 Discussion

In the work described here, we concentrate only on taking advantage of the information that is readily available to us. Ideally, we would like to be able to use the IS asymmetries directly as features, however, without any means of automatically annotating new text with these categories, this is impossible. Our experiments were designed to test, whether we can achieve an improvement in the generation of German text, *without* a fully labelled corpus, using the insight that at least some IS categories correspond to morphosyntactic characteristics that *can* be easily identified. We do not claim to go beyond this level to the point where true IS labels would be used, rather we attempt to provide a crude approximation of IS using only morphosyntactic information. To be able to fully automatically annotate text with IS labels, one would need to supplement the morphosyntactic features

---

[9]The difference may also be due to the fewer features used in the second experiment. However, this emphasises, that the asymmetries gleaned from syntactic information alone are not strong enough to be able to determine the prevailing order of constituents. When we take the IS labels into account, we are honing in on a particular subset of interesting syntactic asymmetries.

with information about anaphora resolution, world knowledge, ontologies, and possibly even build dynamic discourse representations.

We would also like to emphasise that we are only looking at one sentence at a time. Of course, there are other inter-sentential factors (not relying on external resources) that play a role in choosing the optimal string realisation, for example parallelism or the position of the sentence in the paragraph or text. Given that we only looked at IS factors within a sentence, we think that such a significant improvement in BLEU and exact match scores is very encouraging. In future work, we will look at what information can be automatically acquired to help generation ranking based on more than one sentence.

While the experiments presented this paper are limited to a German realisation ranking system, there is nothing in the methodology that precludes it from being applied to another language. The IS annotation scheme is language-independent, and so all one needs to be able to apply this to another language is a corpus annotated with IS categories.

We extracted our IS asymmetry patterns from a small corpus of spoken news items. This corpus contains text of a similar domain to the TIGER treebank. Further experiments are required to determine how domain specific the asymmetries are.

Much related work on incorporating information status (or information structure) into language generation has been on spoken text, since information structure is often encoded by means of prosody. In a limited domain setting, Prevost (1996) describes a two-tiered information structure representation. During the high level planning stage of generation, using a small knowledge base, elements in the discourse are automatically marked as new or given. Contrast and focus are also assigned automatically. These markings influence the final string generated. We are focusing on a broad-coverage system, and do not use any external world-knowledge resources. Van Deemter and Odijk (1997) annotate the syntactic component from which they are generating with information about givenness. This information is determined by detecting contradictions and parallel sentences. Pulman (1997) also uses information about parallelism to predict word order. In contrast, we only look at one sentence when we approximate information status, future work will look at cross sentential factors. Endriss

and Klabunde (2000) describe a sentence planner for German that annotates the propositional input with discourse-related features in order to determine the focus, and thus influence word order and accentuation. Their system, again, is domain-specific (generating monologue describing a film plot) and requires the existence of a knowledge base. The same holds for Yampolska (2007), who presents suggestions for generating information structure in Russian and Ukrainian football reports, using rules to determine parallel structures for the placement of contrastive accent, following similar work by Theune (1997). While our paper does not address the generation of speech / accentuation, it is of course conceivable to employ the IS annotated radio news corpus from which we derived the label asymmetries (and which also exists in a spoken and prosodically annotated version) in a similar task of learning the correlations between IS labels and pitch accents. Finally, Bresnan et al. (2007) present work on predicting the dative alternation in English using 14 features relating to information status which were manually annotated in their corpus. In our work, we manually annotate a small corpus in order to learn generalisations. From these we learn features that approximate the generalisations, enabling us to apply them to large amounts of unseen data without further manual annotation.

8 Conclusions

In this paper we presented a novel method of including IS into the task of generation ranking. Since automatic annotation of IS labels themselves is not currently possible, we approximate the IS categories by their syntactic characteristics. By calculating strong asymmetries between pairs of IS labels, and establishing the most frequent syntactic characteristics of these asymmetries, we designed a new set of features for a log-linear ranking model. In comparison to a baseline model, we achieve statistically significant improvement in BLEU score. We showed that these improvements were not only due to the effect of purely syntactic asymmetries, but that the IS asymmetries were what drove the improved model.

References

Betty J. Birner. 1994. Information Status and Word Order: an Analysis of English Inversion. *Language*, 70(2):233–259.

Joan Bresnan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the Dative Alternation. *Cognitive Foundations of Interpretation*, pages 69–94.

Aoife Cahill, Martin Forst, and Christian Rohrer. 2007. Stochastic Realisation Ranking for a Free Word Order Language. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 17–24, Saarbrücken, Germany. DFKI GmbH.

Herbert H. Clark and Catherine R. Marshall. 1981. Definite Reference and Mutual Knowledge. In Aravind Joshi, Bonnie Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press.

Kees van Deemter and Jan Odijk. 1997. Context Modeling and the Generation of Spoken Discourse. *Speech Communication*, 21(1-2):101–121.

Cornelia Endriss and Ralf Klabunde. 2000. Planning Word-Order Dependent Focus Assignments. In *Proceedings of the First International Conference on Natural Language Generation (INLG)*, pages 156–162, Morristown, NJ. Association for Computational Linguistics.

Martin Forst. 2007. *Disambiguation for a Linguistically Precise German Parser*. Ph.D. thesis, University of Stuttgart. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), Vol. 13(3).

John A. Hawkins. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. Croom Helm, London.

Ron Kaplan and Joan Bresnan. 1982. Lexical Functional Grammar, a Formal System for Grammatical Representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. MIT Press, Cambridge, MA.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In Dekang Lin and Dekai Wu, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 388–395, Barcelona. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, pages 311–318, Philadelphia, PA.

Scott Prevost. 1996. An Information Structural Approach to Spoken Language Generation. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL 1996)*, pages 294–301, Morristown, NJ.

Ellen F. Prince. 1981. Toward a Taxonomy of Given-New Information. In P. Cole, editor, *Radical Pragmatics*, pages 233–255. Academic Press, New York.

Ellen F. Prince. 1992. The ZPG Letter: Subjects, Definiteness and Information Status. In W. C. Mann and S. A. Thompson, editors, *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, pages 295–325. Benjamins, Amsterdam.

Stephen G. Pulman. 1997. Higher Order Unification and the Interpretation of Focus. *Linguistics and Philosophy*, 20:73–115.

Arndt Riester. 2008a. A Semantic Explication of 'Information Status' and the Underspecification of the Recipients' Knowledge. In Atle Grønn, editor, *Proceedings of* Sinn und Bedeutung *12*, University of Oslo.

Arndt Riester. 2008b. *The Components of Focus and their Use in Annotating Information Structure*. Ph.D. thesis, University of Stuttgart. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), Vol. 14(2).

Christian Rohrer and Martin Forst. 2006. Improving Coverage and Parsing Quality of a Large-Scale LFG for German. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2006)*, Genoa, Italy.

Rob van der Sandt. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9:333–377.

Mariët Theune. 1997. Goalgetter: Predicting Contrastive Accent in Data-to-Speech Generation. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL/EACL 1997)*, pages 519–521, Madrid. Student paper.

Nadiya Yampolska. 2007. Information Structure in Natural Language Generation: an Account for East-Slavic Languages. Term paper. Universität des Saarlandes.