Information Structure Annotation:
Adapting Potsdam SFB 632 Information Structure Guidelines to Narrative Data Containing Dialogues

Varya Gracheva and Sanghoun Song
{gracheva, sanghoun}@uw.edu
(Dept. of Lingusitics, Univ. of Washington)

Contents

I. INTRODUCTION

The present guidelines were developed for annotation of the information structure in the narrative text, specifically *The Little Prince*. The goal of this annotation project is to annotate a parallel text (existing original and translations of *The Little Prince*) in order to capture the information structure phenomena happening cross-linguistically. Based on the existing Potsdam SFB 632 Information Structure Guidelines, the present guidelines can be seen as their expansion, aimed at analyzing the peculiarities of the narrative text, such as the co-occurrence of the running text with the question-answer sequences in this kind of data. These present guidelines can hopefully aid in future annotation tasks of similar types of texts. The Leipzig Glossing guidelines were also used in this project, however also somewhat modified to better suit our project goals. Most of examples in these guidelines are provided for Russian, Korean, Spanish, and English data, but can be used to aid in the annotation of the data in other languages. This is still work in progress; more examples from our annotation data and explanations will be added to this document later.

II. GLOSS

1. Segmentation: Hyphens vs. Colons

Our gloss annotation schema is based off of the Leipzig Glossing Rules and the POTSDAM SFB 632 Annotation Guidelines with some adaptations included to better suit our annotation needs. The similarities and differences between these two guidelines and our guidelines are explained below.

(a). Using hyphens

Following POTSDAM SFB 632 and Leipzig annotation guidelines, hyphens are used for segmentation. Below is an example of a glossed heavily-inflected verb:

TXT	рассказывалось
MORPH	рассказыва-л-о-сь
GLOSS	tell-IPFV.PST-SG.N-REFL

(b). Using colons

When there is no segmentation—e.g., in uninflected forms—a colon is used instead of a hyphen as in:

TXT	принц
MORPH	принц
GLOSS	prince:M.SG.NOM

or

TXT	лет
MORPH	лет
GLOSS	year:PL.GEN

2. Different number of segments in the Gloss Tier

(a). one word

When several words are translated as one word, the cells are merged in the Gloss, as in example below:

TXT	стало	быть
GLOSS	Hence	

(b). more than one word

Following Leipzig Glossing Rules, if a word's translation consists of two or more words, an underscore is used as a separator in the Gloss tier:

TXT	совсем
GLOSS	at_all

3. Glossing Punctuation Marks

All punctuation marks, following POTSDAM SFB 632 guidelines, are present in separate cells in the TXT layer and are absent in GLOSS and MORPH layers:

TXT	-	спросил	Маленький	принц	.
GLOSS		ask-PFV.PST.3SG.M	little-M.SG.NOM	prince:M.SG.NOM	

4. New Tags in Gloss

The original tag set has been supplemented in order to more robustly capture language specific information:

TAG	**CONSTRUCTION**	**EXAMPLE**
IMP	Imperatives	Spanish *dibuja*
INT	Interrogative elements	Russian particle *li* (when it's only interrogative and does not mark focus)
COND	Conditional elements	Russian particle *бы*
CONT	Contrastive elements	Russian particles *zhe, -to, ved',* when used contrastively (i.e. sentential vs. smaller constituent attachment, e.g. non-contrastive indefinite *chto-to*)
FOC	Focus elements	Russian particle *li* (interrogative particle *li* when it marks focus on the preceding constituent)
TOP	Topic elements	Topic markers (-(*n*)*un* in Korean)
PART		Russian particles *-ka, -nibud, za*
REFL	Reflexive elements	Russian suffix *–sja*
PRET	Preterits	Spanish *fui*, *supe*

5. Pronouns and Clitics

Pronouns and clitics are glossed as person, number, gender, and case.

TXT	he
GLOSS	3SG.M.NOM

III. NP_TYPE

1. NP Tags List

TAG	Categories	Example(s)
All	Universal	*All*
Alt	Alternative	*another, other*
Any	NPI	*any*
Bare	Bare	*e.g. flowers*
Class	Classifier	*Classifiers in Chinese*
Cleft	Clefting	*Clefting clauses in Korean*
Def	Definite	*the*
Dem	Demonstrative	*these*
Dist	Distributive	*each, every*
Ind	Indefinite	*any, some, etc.*
Kind	Kindness	*such as*
Mul	Multal	*much, many (many years ago)*
Neg	Negative	*no, neither*
Num	Numeral	*one, twelve*
Ord	Ordinal	*first, tenth*
Pau	Paucal	*little, few*
Pro	Pronoun	*you, he, were*
Refl	Reflexive	*self*
Sup	Superlative	*most*
Wh	Wh-words	*what, which*
CL	Clitics	*lo, la, les, se* in Spanish
Un	Uniquitive	e.g. *the only*

2. Annotating Maximal NP(DP) instead of a noun only

Maximal NPs (DPs) are annotated at the level of NP_TYPE. For example, the entire NP/DP "the little prince" should be marked as *def*, as opposed to only marking the noun "prince".

3. Annotating Two Determiners in One NP(DP)

If there are two determiners within one NP, a semicolon is used to separate the tags describing determiners, as in the example below:

TXT	all	the	books
NP_TYPE	all;def		

Tags "all;def" are used to show that there are a universal and a definite determiners present in the NP "all the books".

IV. DROPPED

A hash mark "#" is used to indicate a presence of a dropped element (i.e. pronouns, subjects, topics). It is inserted into its own cell in the text tier in the position where it would most likely have appeared had it not been dropped. Below we distinguish the lack of a pronoun/subject/topic from the cases where it is dropped.

(a). Elements treated as NOT dropped:

- Absent pronouns in the Imperatives (Ex. *Draw me a sheep*.)
- Absent nouns following adjectives/determiners that can themselves be treated as nouns. We test the "noun nature" of the element in question by checking whether (1). if the dropped element would not be actually dropped, it enters in an agreement relationship with the accompanying verb and (2). it can be used on its own (without a noun) in another sentence. If the adjective/determiner element satisfies both of these conditions we treat it as a noun, without a dropped element before or after it. Examples include *drugoi* in Russian and *otro* in Spanish (both roughly translated as "other" in English). In the sentence below "other" is treated as a noun, without a dropped element before or after it, and therefore the DROPPED tier remains empty:

TXT	Нарисуй	другого	.
GLOSS	Draw-IMP.2SG	other-SG.M.ACC	
DROPPED			

Further evidence is found in Russian sentences like *Другие придут нам на смену/ Another-PL will come to us on change / 'Others will come to replace us'*, in which "another-PL" exists on its own, without a noun.

In the English example below determiner "this" functions as a noun/pronoun and adjective "old" functions as a predicate. As in the example above, there is no dropped element in this sentence although it is possible for a noun to grammatically appear after 'this':

TXT	This	is	too	old	.
GLOSS	This	is	too	old	
DROPPED					

- **PRO** subjects in the tenseless clauses are not considered to be dropped elements:

  Ex. In order for *her* to read this book she needs to buy it.
  This food makes *me* hungry.
  It is hard for *me* to read without my glasses.

  In the above, 'for her' in the first line and 'for me' in the third line can disappear even in English, which we do not consider dropped elements from the given sentence. Moreover, in the corresponding Korean and Chinese translations of the examples in the above, the italic pronouns can be missing. In that case, we do not assume that the missing pronouns are dropped elements.

(b). Elements treated as dropped:

- The **pro** subjects in tensed clauses, on the other hand, are considered to be dropped in our annotation. In the Russian example below there is a dropped element in sentence-initial position, annotated in more detail below:

[TXT]	#	Протёр	глаза	.
[GLOSS]		rub:PFV.PST.SG.M	Eye-PL.ACC	
[DROPPED_WORD]	I			
[DROPPED_FEAT]	1.SG.NOM			
[DROPPED_IDX]	2.13.0			

The dropped element indicated by "#" in the TXT layer, refers to the word "I" in the DROPPED_WORD layer. In the latter layer we use English gloss for the content of the dropped word in our annotation. This element has features 1SG.NOM, annotated on the DROPPED-FEAT layer. Its antecedent can be found in position indexed 2.13.0, which refers to chapter 2, sentence 13, word 0 (described in more detail in the section below).

V. DROPPED_IDX

This layer refers to the index of the antecedent of the dropped element. It shows the position of the word within a chapter and a sentence.

(a). Absent Antecedent

If there is no antecedent for a dropped element, then we leave DROPPED_IDX as empty, for example in 12.6 in *The Little Prince*, there is no antecedent for dropped "I", because it was not mentioned in the previous sentences and therefore the layer DROPPED_IDX remains empty:

[TXT]	#	Пью	,
[GLOSS]		drink-IPFV.PRS.1SG	
[DROPPED_WORD]	I		
[DROPPED_FEAT]	1SG.NOM		
[DROPPED_IDX]			

(b). Antecedent consisting of several words

If the index for a dropped element has to reference two or more words (i.e. *Little Prince*, the relevant indices are separated with a semicolon:

[TXT]	И	#	прибавил	не	без	грусти	:
[GLOSS]	and		add-PFV.PSG.SG.M	not	without	sadness-SG.F.GEN	
[DROPPED_WORD]		Little_Prince					
[DROPPED_FEAT]		2SG.M.NOM					
[DROPPED_IDX]		3.48.2;3.48.3					

VI. INFORMATION STRUCTURE (IS): GENERAL REMARKS

1. New Chapter

The beginning of every chapter is treated as continuation of the previous chapter. This is particularly important for annotating Information Status, when choosing between *giv-active*, *giv-inactive*, *new* and *acc* on the INFOSTAT layer (these tags adapted from the POTSDAM SFB 632 guidelines will be discussed in more detail later). For example, the element in question in the first sentence of a chapter is annotated as *giv-active*, if it has an antecedent in the last sentence of the previous chapter.

2. Annotating NP (DP) for IS

NPs are the only constituents that are annotated on the INFOSTAT layer. As with the other layers, annotation should cover the maximal NP. For example, the entire NP "The Little Prince," should be marked as *giv*, as opposed to annotating just the noun.

3. Dialogue and Author's Narration in Running Text: Outer Frame and Inner Frame

The Little Prince annotated for this project is an example of a running text, in which "written" and "spoken" text styles co-occur. There are author's narrative (treated as written text) and dialogues between characters (treated somewhat as spoken text). In order to accurately annotate IS on these different levels, we have introduced a differentiation between an Outer Frame and an Inner Frame. The Outer Frame (OF) refers to the author's narration, the frame of the written (or also referred to as running) text. The Inner Frame (IF) refers to situated dialogue between characters. Below is annotation of a sentence "I am drinking, -replied the tippler", which is a response to the Little Prince's question "What are you doing?"

Example: *"Пью," - мрачно ответил пьяница.*

TXT	-	#	Пью	,	-	мрачно	ответил	пьяница	.
GLOSS			drink-IPFV.PRS.1SG			gloomily	answer-PFV.PST.SG.M	tippler-SG.M.NOM	
OF-INFOSTAT								giv-active	
OF-TOPIC									
OF-FOCUS			nf-unsol			nf-unsol			
IF-INFOSTAT									
IF-TOPIC			nf-sol						
IF-FOCUS									

The OF and IF are annotated separately as follows:

a) **Outer Frame**

The quotation marks set off their content from the rest of the text as *nf-(un)sol*, and, in this case, all remaining text is background information. In the example above the content within the quotation marks (i.e. "I am drinking") is treated as *nf-unsol*. We treat it as "unsolicited" response on this level, because even though on the question-answer sequence level it is a solicited response to a question, on the running text level it is an unsolicited element, though still focused as a part of utterance that can be paraphrased as "The tippler gloomily answered that he **is drinking**". Although content set off with quotation marks is always focused within the OF, additional content may be annotated as *background, topic,* or *focus*, depending on the context. For more information on how these distinctions are made see the TOPIC section.

b). **Inner Frame**

The phrase within the quotation-marks is analyzed as a continuation of the preceding dialogue, in this case *nf-sol* (as opposed to *nf-unsol* on the OF level), since the utterance is a response to the Little Prince's question "What are you doing?" None of the following phrase is annotated on this layer, because it is exclusively part of the author's narration.

4. INFOSTAT Layer in IF

The pronoun elements referring to the dialogue characters (such as '*you*' or '*I*') within the dialogue are going to be annotated for INFOSTAT only on the IF layer. Starting from the first dialogue between characters, the pronouns *you* and *I* are labeled as ***acc-sit*** (*accessible-situative*). Even though their referents have not yet been overtly mentioned on the IF level (i.e. within the dialogue), they are accessible from what is referred to as "situative context" in POSTSDAM SFB 632 project, and are now a part of the discourse situation (POTSDAM SFB 632, p. 156-157).

The rest of the NPs mentioned in the dialogue between characters are also going to be annotated as ***new*** if the character introduces the corresponding referents, or ***acc-sit*** if they are already a part of the discourse situation.
An example of the annotated first phrase in the dialogue between Little Prince and the author, exhibiting both the ***acc-sit*** and ***new*** elements, is below:

TXT	Нарисуй	мне	барашка	…
GLOSS	draw-PFV.IMP.2SG	1SG-DAT	sheep-SG.M.ACC	
OF-INFOSTAT				
IF-INFOSTAT		acc-sit	new	

Since this is the first interaction between the two characters and first mention of the sheep, pronoun '1SG-DAT' is annotated as ***acc-sit*** and noun 'sheep' is annotated as ***new***.
Below is an example of a later interaction between two characters (the same sentence is repeated again), in which both the pronoun '1SG-DAT' and noun 'sheep' are therefore annotated as ***acc-sit***, because both of them are now a part of the shared discourse between two characters:

TXT	Нарисуй	мне	барашка	…
GLOSS	draw-PFV.IMP.2SG	1SG-DAT	sheep-SG.M.ACC	
OF-INFOSTAT				
IF-INFOSTAT		acc-sit	acc-sit	

VII. TOPIC

More to be added to this section later, including the tests for [ab(outness)] topics. This is necessary to avoid annotating too many elements as [ab] topics. As of now, we are assuming the POSTDAM SFB 632 guidelines for annotating our data.

VIII. FOCUS

Following POTSDAM SFB 632 guidelines, we treat *The Little Prince* as a running text (more precisely, narrative), annotating the focus elements in author's narration, i.e. on the outer frame level, as "*unsolicited new-information focus*" (*nf-unsol*) (POTSDAM SFB 632, p. 176-177), as they carry forward the discourse, but do not represent solicited information. The process of determining the focused elements is also adapted from POTSDAM SFB 632, i.e. we assume that for each sentence in our text there exists a preceding implicit question (POTSDAM SFB 632, p. 176). We formulate this question to the best of our ability and determine the answer, which refers to the new information and therefore is a focused element.

Annotating the focus elements in the dialogues within our text, i.e. on the inner frame level, is different. We analyze the inner frame as a spoken speech, thus the new information can be both unsolicited (e.g. a question), annotated as *nf-unsol*, and solicited (e.g. an answer), annotated as *nf-sol*.

TXT	Нарисуй	мне	барашка	…
GLOSS	draw-PFV.IMP.2SG	1SG-DAT	sheep-SG.M.ACC	
OF-INFOSTAT				
OF-TOPIC				
OF-FOCUS	nf-sol			
IF-INFOSTAT		acc-sit	acc-sit	
IF-TOPIC				
IF-FOCUS	nf-sol			

IX. CONTRAST

As in POTSDAM SFB 632 guidelines, we annotate contrast in our project. We have introduced a separate layer for contrast to address the contrastive focus and contrastive topic found in *The Little Prince*. Below are the three different tags for the CONTRAST layer that we have introduced:

CT	Contrastive Topic
CF	Contrastive Focus
C?	When in doubt

More examples will be added later to this section.

X. INFORMATION STRUCTURE: MORE SPECIFIC ISSUES

1). Yes-No questions

We annotate Yes-No questions as all-focus sentences:

TXT	Are	you	happy	?
FOCUS	nf-unsol			

2). *Wh*-expressions

It is traditionally assumed that in the interrogative clauses the interrogative pronoun universally bears the focus function (Bjerre 2011). We expanded on this treatment of interrogative pronouns for our annotation, introducing the distinction between narrow and all-sentence focus depending on the type of frame in the running text it is found in (i.e. Inner Frame vs. Outer Frame). *Wh*-exclamations are treated differently from the *wh*-questions.

(a). *Wh*-questions

Wh-questions are annotated with *wh*-words as focused elements. In the example below "where" is annotated as focused element, while the rest of the phrase is a background:

TXT	Where	is	He	?
FOCUS	nf-unsol			

An example, where the focus can be different depending on the layer annotated (i.e. OF vs. IF), is below:

TXT	-	Что	это	за	штука	?
GLOSS		what	this	PART	thing-SG.F.NOM	
OF-FOCUS		nf-unsol				
IF-FOCUS		nf-unsol				

The *wh*-word "what" is the only focused element on the IF level, while the entire sentence "What is this thing?" is focused on the OF level.

(b). *Wh*-exclamations

Wh-questions are annotated differently from *wh*-exclamations, in which the *wh*-word is not necessarily the only focused element. An example of *wh*-exclamation is a sentence "What a beauty!" which is annotated as an all-focus sentence.

3). Imperatives

We treat Imperatives somewhat similar to the other constructions.

The Imperatives that are similar to the "all-new/event" sentences (POTSDAM SFB 632, 163) are annotated as all-focus sentences:

TXT	Watch	out	for	the	baobabs	!
FOCUS	nf-unsol					

If the addressee is present in the Imperative construction, we treat the addressee as a background, since 1) it does not appear to be what the sentence is about (hence, it is not an aboutness topic), 2) it does not set the frame for the situation (hence, it is not a frame-setting topic), and, finally, 3) it does not appear to carry the discourse forward and can be omitted (hence, it is not focus):

TXT	Children,	watch	out	for	the	baobabs	!
TOPIC							
FOCUS		nf-unsol					

The other Imperatives that are not all-new/event sentences, are treated as sentences that can have topics/foci/contrasts. For example, in the sentence below we have "blue" annotated as focused/contrastive element, while the rest of the sentence is a background:

TXT	Go	to	the	blue	one	!
FOCUS				nf-unsol		
CONTRAST				cf		

4). Complex Sentences

Complex sentences are analyzed as separate clauses, each with its own topic and focus (if applicable). Thus complex sentences may have multiple foci and topics. The exceptions to treating complex sentences as separate independent clauses are:

a. Direct Speech

Ex. *What are you doing? -asked the Little Prince*
("*What are you doing?"* = nf-unsol, "*asked the Little Prince*" = bg, on the OF level)

Direct speech in running text can have various representations (see examples below for variations in different languages):

"----" S(top) V	S has topic marker (Korean, Chinese)
"----" S(bg) V	S should be marked as bg or topic, depending on the context (Russian, Spanish)
"----" V(bg) S	V should be marked as bg or topic, depending on the context (Russian, Spanish)
# bg V	The noun can be dropped. Only verb remains. (Spanish)
#	Entire phrase following direct speech, describing who is talking, can be dropped (English, Russian, Spanish).

Below is an example of annotated direct speech:

TXT	Отчего	же	тебе	совестно?	спросил	Маленький	принц,	ему	очень	хотелось	помочь	бедняге	
[MORPH]	Отчего	же	теб-е	совестно	спроси-л	Маленький	принц	е-му	очень	хотел-о-сь	помочь	бедняг-е	
[GLOSS]	why	PA	2SG-D	ashamed	ask-PFV.PST.3M.SG	little-SG.M.NOM	prince:SG.M.NOM	3SG.M-DAT	very	want-IPFV.PS	help	poor_fellow-SG.M.DAT	
[NP_TYPE]													
[DROPPED_WORD]													
[DROPPED_FEAT]													
[DROPPED_IDX]													
[0F-INFOSTAT]			acc			giv-inactive		giv-active				giv-active	
[OF-TOPIC]						ab		ab					
[OF-FOCUS]	nf-unsol								nf-unsol				
[OF-CONTRAST]													
[IF-INFOSTAT]													
[IF-TOPIC]			ab										
[IF-FOCUS]	nf-unsol												
[IF-CONTRAST]													

e. Temporal PPs

When the clause is adjoined to the whole sentence, it should be treated as a part of this sentence, not as a separate clause. For example, in 12.5 below we annotate the entire temporal PP "when he appeared on this planet" as frame-setting topic:

TXT	When	he	appeared	on	this	planet	,	the	tipple r	was	…
TOPIC	fs										

f. Temporal PPs: Contrastive Focus vs. Contrastive Topic

In some cases the temporal PPs in the beginning of the sentence are annotated as focus, for example in the example below. "not soon" below is annotated as a contrastive focus because it a). brings new information and b). is contrastive:

TXT	Не	скоро	я	понял	**,**	откуда	он	явился	**.**
GLOSS	Not	soon	1SG.NOM	understand-PFV.PST.SG.M		where_from	3SG.M	appear-PFV.PST.SG.M	
FOCUS	nf-unsol_1					*nf-unsol_2			
CONTRAST	cf								

g. Relative Clauses:

In our analysis of relative clauses we have examined 3 possible situations, each with different implications for the topic and focus annotation. Even though in relative clauses the relative pronoun is usually treated as bearing the topic function (Bjerre 2011), we somewhat deviate from this traditional approach for our information structure annotation purposes, usually treating the relative pronoun as a part of focus.

(1). Situation 1

Question: *Whom did you meet?*
Answer: *I met a girl who has a good house.*

TXT	I	met	a	girl	who	has	a	good	house	.
TOPIC										
FOCUS			nf-sol							

(according to POTSDAM SFB 632 guidelines, p. 173)

(2). Situation 2

Question: *What kind of girl did you meet?*
Answer: *I met a girl who has a good house.*

TXT	I	met	a	girl	who	has	a	good	house	.
TOPIC			ab							
FOCUS					nf-unsol					

(according to POTSDAM SFB 632 guidelines, p. 163 (on [ab] topic) and p. 173 (on [nf] focus))

(3). Situation 3

Question: *What happened yesterday?*
Answer: *I met a girl who has a good house.*

TXT	I	met	a	girl	who	has	a	good	house	.
TOPIC										
FOCUS	nf-unsol									

(according to POTSDAM SFB 632 guidelines, p. 178)

5). Copula Constructions

If a verb in the copular construction does not carry semantic content, it will be treated as background.

a). Presentational Copula Construction

Copular verb "be" is background in this case, not marked for Topic or Focus.

TXT	She	is	kind	.
TOPIC	ab			
FOCUS			nf-unsol	

b). Identity Copula Construction

Copular verb "be" is background in this case, not marked for Topic or Focus.

TXT	She	is	Mary	.
TOPIC	ab			
FOCUS			nf-unsol	

c). Existential Copular Construction

A context for this construction is below:
Question: *Is she in the garden?*
Answer: *She is.*

In this and similar contexts, a copular verb "be" can be annotated as Focus or a part of Focus.

TXT	She	is	.
TOPIC			
FOCUS	nf-sol		

If the same example includes "yes", we annotate this sentence as having two foci, one of which is the phrase including "is":

TXT	Yes,	she	is	.
TOPIC				
FOCUS	nf-unsol	nf-unsol		

However, in the context, where *"is*" is not bringing any new information, such as in the context below:

Question: *Where is she?*
Answer: *She is in the garden.*

the copular verb "*be*" is annotated as background, while "*in the garden*" is annotated as focus:

TXT	She	is	in	the	garden	.
TOPIC						
FOCUS			nf-sol			

6). Primary and Secondary Focus/Discontinuous Focus

Following the POTSDAM SFB 632 annotation guidelines, we allow for presence of primary and secondary foci in the same sentence, marked nf_1 and nf_2 respectively. We also added asterisk "*" to annotate a primary focus, the main characteristic of which is that it cannot be dropped, because it would cause the sentence to lose too much information.

In the example below, *physically* is a continuation of the same focus, starting with *how*, referred to as "discontinuous focus domain" in the POTSDAM SFB 632 guidelines (pp. 175-176).

TXT	How	was	he	doing	physically	?
FOCUS	*nf_1				nf_1	

In the example below the clause within the quotation marks is treated as the primary focus:

TXT	"	I	am	drinking	"	-	replied	the	tippler	with	a	lugubrious	air	.
FOCUS		*nf_1								nf_2				

7). ADVERBIALS: Focus vs. Topic vs. Background

If the element in question (adverbial construction referring to as "denoting domains against which the subsequently reported fact is to be evaluated", per POTSDAM SFB 632, p. 168), is placed in the end of the sentence, we treat it as contrastive focus.

TXT	He	is	doing	well	physically	.
FOCUS					nf-unsol	
CONTRAST					cf	

If this adverbial is fronted, we treat it as a frame-setting topic/contrastive topic:

TXT	Physically,	he	is	doing	well	.
TOPIC	fs				nf-unsol	
CONTRAST	ct					

At present this analysis applies to English only. Its other limitation is that this analysis is based on the form rather than on the meaning. Next step is to figure out how this analysis applies across other languages, based on the meaning of the adverb as opposed to its form.

8). "There" vs. "here" vs. "this is" constructions

If someone is handing a drawing, saying, "Here is the drawing", we treat *here* as focus and the rest of the sentence as a background:

TXT	Here	is	the	drawing	.
FOCUS	nf-unsol				

If someone mentions that there exists a copy of the drawing, we treat "*a copy of the drawing*" as a focus, and the rest of the sentence as a background:

TXT	There	is	a	copy	of	the	drawing	.
FOCUS			nf-unsol					

If someone points out that the object in question is not a drawing, we treat "not a drawing" as focus, ad the rest of the sentence as a background:

TXT	This	is	not	a	drawing	.
FOCUS			nf-unsol			

References

Avgustinova, Tania and Yi Zhang. 2010. Conversion of a Russian dependency treebank into HPSG derivations. In *Proceedings of the 9th International Workshop on Treebanks and Linguistic Theories (TLT'9)*, Tartu, Estonia.

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar Customization. *Research on Language & Computation*, 8(1):23–72.

Bjerre, Anne (2011). Topic and focus in local subject extractions in Danish. Presented at the 18th International Conference on Head-Driven Phrase Structure Grammar, August 23, 2011.

Bond, Francis, Sanae Fujita , Chikara Hashimoto , Kaname Kasahara , Shigeko Nariyama , Eric Nichols , Akira Ohtani , Takaaki Tanaka , and Shigeaki Amano. 2004. The Hinoki Treebank: A Treebank for Text Understanding. In *Proceedings the 1st IJCNLP*, pages 158–167.

Bouma, Gerlof, Lilja Øvrelid, and Jonas Kuhn. 2010. Towards a Large Parallel Corpus of Cleft Constructions. In *Proceedings of LREC*, pages 3585–3592.

Büring, Jeanette K. 1999. Topic. In Bosch, Peter and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 142–165. Cambridge University Press, Cambridge.

Büring, Jeanette K. 1999. Topic. In Bosch, Peter and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 142–165. Cambridge University Press, Cambridge.

Calhoun, Sasha, Malvina Nissim, Mark Steedman, and Jason Brenier. 2005. A Framework for Annotating Information Structure in Discourse. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 45–52. Association for Computational Linguistics.

Copestake, Ann and Dan Flickinger. 2000. An Open-Source Grammar Development Environment and Broad-Coverage English Grammar using HPSG. In *Proceedings of the 2nd conference on Language Resources and Evaluation*, Athens, Greece.

Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology and Department of Linguistics of the University of Leipzig. *Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Retrieved from http://www.eva.mpg.de/lingua/resources/glossing-rules.php

Dipper, Stefanie, Michael Goetze, and Stavros Skopeteas. 2007. *Information Structure in Cross-linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*. Universitätsverlag Potsdam.

Engdahl, Elisabet and Enric Vallduví. 1996. Information Packaging in HPSG. *Edinburgh Working Papers in Cognitive Science*, 12:1–32.

Gundel, Jeanette K. 1999. On Different Kinds of Focus. In Bosch, Peter and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 293–305. Cambridge University Press, Cambridge.

Johansson, Mats. 2001. Clefts in Contrast: a Contrastive Study of it Clefts and wh Clefts in English and Swedish Texts and Translations. *Linguistics*, 39(3):547–582.

Komagata, Nobo N. 1999. *A Computational Analysis of Information Structure Using Parallel Expository Texts in English and Japanese*. Ph.D. thesis, University of Pennsylvania.

Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, UK.

Marimon, Montserrat, Núria Bel, Sergio Espeja, and Natalia Seghezzi. 2007. The Spanish Resource Grammar: pre-processing strategy and lexical acquisition. In *Proceedings of the Workshop on Deep Linguistic Processing*, pages 105–111. Association for Computational Linguistics.

Och, Franz J. and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, 29(1): 19-51.

Oepen, Stephan, Daniel Flickinger, Kristina Toutanova, and Christopher D. Manning. 2004. LinGO Redwoods: A Rich and Dynamic Treebank for HPSG. Journal of Research on Language and Computation, 2(4): 575–596.

Oepen, Stephan, Erik Velldal, Jan T. Lønning, Paul Meurer, Victoria Rosén, and Dan Flickinger. 2007. Towards Hybrid Quality-Oriented Machine Translation. – On linguistics and probabilities in MT –. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation*, Skvde, Sweden.

Song, Sanghoun, Jong-Bok Kim, Francis Bond, and Jaehyung Yang. 2010. Development of the Korean Resource Grammar: Towards Grammar Customization. In *Proceedings of the 8th Workshop on Asian Language Resources*, pages 144–152. Beijing, China.