Towards Interoperability for the Penn Discourse Treebank

Nancy Ide
Department of Computer Science
Vassar College
ide@cs.vassar.edu

Rashmi Prasad
Institute for Research in Cognitive Science
University of Pennsylvania
rjprasad@seas.upenn.edu

Aravind Joshi
Institute for Research in Cognitive Science
University of Pennsylvania
joshi@seas.upenn.edu

Abstract

The recent proliferation of diverse types of linguistically annotated schemes coded in different representation formats has led to efforts to make annotations interoperable, so that they can be effectively used towards empirical NL research. We have rendered the Penn Discourse Treebank (PDTB) annotation scheme in an abstract syntax following a formal generalized annotation scheme methodology, that allows meaning-preserving mappings to any other scheme. As an example, we show the mapping of the PDTB abstract syntax to a representation in the GrAF format.

1 Introduction

The last decade has seen a proliferation of linguistically annotated corpora coding many phenomena in support of empirical natural language research – both computational and theoretical. Because the annotated phenomena and annotation representations vary widely across different schemes, there is a need for making them compatible with each other, to ensure effective merging, comparison and manipulation with common software.

Ensuring compatibility is even more necessary when different types of annotations are done on the same source text, for example the *Wall Street Journal* (*WSJ*) corpus. Multi-level annotations on the *WSJ* include part of speech tagging, syntactic constituency, coreference, semantic role labeling, events, and discourse relations. In many cases, empirical natural language research using the *WSJ* would benefit immensely from using information from multiple layers of annotation, but in order to allow for this, it is imperative to ensure the efficient interoperability of the annotations.

In Bunt (2010), an annotation scheme design methodology is proposed, providing a formal specification of a representation format as a rendering of conceptual structures defined by an *abstract syntax*. It ensures that every representation format is convertible through a meaning-preserving mapping to any other representation format. In Ide and Bunt (2010), Bunt's design methodology is further generalized, and a mapping strategy is defined to convert from an abstract syntax to a representation in GrAF format (Ide and Suderman, 2007). To illustrate the process, Ide and Bunt generate an abstract syntax and apply the mapping strategy to annotation schemes for ISO-TimeML (ISO, 2009), PropBank (Palmer et al., 2005), and FrameNet (Baker et al., 1998).

In this paper, we present an application of this methodology to the annotation scheme of the Penn Discourse Treebank (PDTB) (Prasad et al., 2008), which contains annotations of discourse relations on the *WSJ*. Our goal is to allow for effective combination of PDTB with GrAF renderings of PropBank and other annotations that have been done on all or parts of the *WSJ*, including Penn Treebank (PTB) syntactic annotations. In addition, we hope that this work will feed the development of a standard for annotating discourse relations in ISO project 24617-5 (Semantic Annotation Framework, Part 5: Discourse Relations).

2 The PDTB: Brief Overview

The PDTB (Prasad et al., 2008) provides annotations of discourse relations, along with their arguments, senses and attributions, on the entire PTB-II portion of the *WSJ* corpus, consisting of approximately 1 million words.[1] Some PDTB annotations are illustrated in Exs.(1-5). Discourse relations, such as causal, contrastive, and temporal relations, are triggered by explicit words or phrases (e.g., the underlined expressions in Exs. (1) and

[1] http://www.seas.upenn.edu/˜pdtb. The corpus is distributed via LDC (http://www.ldc.upenn.edu).

(3), or by adjacency. Explicit realizations can occur via grammatically defined *connectives* (Ex. 1), or with other grammatically non-conjunctive expressions called *Alternative lexicalizations* (AltLex) (Ex. 3). The two arguments of a discourse relation are abstract objects (AO) in discourse, such as events, states, and propositions, and are labelled Arg1 (shown in italics) and Arg2 (shown in bold). Between two adjacent sentences not related by an explicit connective or AltLex, an implicit discourse relation can be inferred, when the annotator has to *insert* a connective to express the inferred relation (e.g., the implicit connective *because* inserted in Ex. 2). It is also possible for adjacent sentences to *not* be related by a discourse relation, in particular when the sentences are linked by an entity-based coherence relation (EntRel, shown in Ex. 4), or are not related at all via adjacency (NoRel, shown in Ex. 5). For each discourse relation, a sense (shown in parentheses at the end of examples), drawn from a hierarchical sense classification scheme, is provided for the relation. The attribution (to an agent of the AO assertion, belief, fact, or eventuality) of each discourse relation and each of its two arguments is also annotated, along with the attribution text when it is explicit (e.g., the attribution over Arg1 in Ex. 3).

1. *Big buyers like P&G say there are other spots on the globe, and in India, where the seed could be grown. . . .* But **no one as made a serious effort to transplant the crop**. (Comparison:Concession:Contra-expectation)

2. *Some have raised their cash positions to record levels.* Implicit=because **High cash positions help buffer a fund when the market falls**. (Contingency:Cause:Reason)

3. *But a strong level of investor withdrawal is much more unlikely this time around,* fund managers said. **A major reason is that investors already have sharply scaled back their purchases of stock funds since Black Monday**. (Contingency:Cause:Reason)

4. *Pierre Vinken, . . ., will join the board as a nonexecutive director Nov. 29.* EntRel **Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group**.

5. *Jacobs is an international engineering and construction concern.* NoRel **Total capital investment at the site could be as much as $400 million,** . . ..

PDTB annotations are stand-off, in that files containing the annotations are physically separate from the source text files. The PDTB annotation scheme and representation are fully described in the manual (PDTB-Group, 2008).

3 A Formalization of PDTB annotations

The current scheme for annotating a *discourse relation entity* in the PDTB includes a list of values, vertically represented in the annotation files. Values also represent *text spans*, as references to the character offsets in the source text file, and the *PTB alignments* of the text spans, as gorn address (Gorn, 1965) references to nodes in their corresponding PTB constituency trees. The full set of features and descriptions of their value assignments is given in Table 6.

The vertical representation of the PDTB annotations can also be converted to a simpler horizontal format, with each line corresponding to one discourse relation.[2] For this work, we have used the horizontal format in which the values for each field are separated by vertical bars; for example:

```
Explicit|258..262|once|1,0,1,2,0|Ot|
Comm|Null|Null|361..377|1,1;1,2;1,3;1,4|
|Temporal.Asynchronous.Succession||||||
202..257|1,0,0;1,0,1,0;1,0,1,1;1,0,1,3|
Inh|Null|Null|Null|||263..282|1,0,1,2,1|
Inh|Null|Null|Null||||
```

The design methodology outlined in Ide and Bunt (2010) consists of a two-phase process: the specification of (1) an **abstract syntax** consisting of a *conceptual inventory* of the elements from which these structures are built up, and *annotation construction rules*, which describe the possible combinations of these elements into annotation structures; and (2) specification of at least one **concrete syntax** providing physical representations for these structures. This methodology has evolved in the context of developing standardized linguistic annotation schemes within ISO TC37 SC4, the foundation of which is the Linguistic Annotation Framework (LAF), (Ide and Romary, 2004); ISO 24612, 2009. LAF defines an *abstract model* for annotations consisting of a directed graph decorated with feature structures that is realized concretely in an XML serialization, the Graph Annotation Format (GrAF), (Ide and Suderman, 2007). GrAF serves as a *pivot* format into which well-formed annotation schemes may be mapped, thus guaranteeing syntactic consistency and completeness for the purposes of comparison, merging, and transduction to other formats.

In the context of ISO work, the abstract syntax for a given annotation type is developed before

[2] A format conversion tool is available from the PDTB Tools site: http://www.seas.upenn.edu/˜pdtb/PDTBAPI/

any concrete realization is specified. However, because the PDTB annotation scheme already exists, we must "reverse engineer" the abstract syntax. The PDTB scheme is a flat structure in which the same information types are repeated multiple times in order to associate them with different annotation elements. For example, values for the same set of features are given for the connective, and the two arguments. In the conceptual inventory, a structure providing this information will be defined once and re-used where necessary.

An abstract syntax for the PDTB annotations includes the following conceptual inventory:

- A set of discourse relations REL = {*explicitRelation, implicitRelation, alternativeLexicalRelation, entityRelation, noRelation*};
- A pair of arguments ARGS = {ARG_1, ARG_2};
- A finite set F of attribution features = {*Source, Type, Polarity, Determinacy*} where *Source* ={Writer*, Other, Arbitrary, Inherited}; *Type* = {Comm (assertion)*, PAtt (belief), Ftv (fact), Ctrl (action), Null}; *Polarity* = {Null*, Negative}; *Determinacy* = {Null*, Indeterminate} (starred values are defaults).
- a semantic class CLASS;
- A connective head HEAD, consisting of the textual rendering of the head of a connective and its semantic class;
- Implicit connective IC, consisting of the textual rendering of the head of a connective and its semantic class;
- Attribution ATTR, consisting of a textual rendering of an explicit attribution, which may be empty;
- Supplementary text SUP, a reference to a span or spans in the text, which may be empty.

The annotation construction rules are the following:

- An **entityRelationEntity** is a pair $\langle a_1, a_2 \rangle \in$ ARGS.
- An **noRelationEntity** is a pair $\langle a_1, a_2 \rangle \in$ ARGS.
- An **explicitRelationEntity** is a triple $\langle EF, arg_1, arg_2 \rangle$, where EF is an explicitConnectiveStructure.
- An **implicitRelationEntity** is a triple $\langle IF, arg_1, arg_2 \rangle$, where IF is an implicitConnectiveStructure.
- An **altLexRelationEntity** is a triple $\langle AF, arg_1, arg_2 \rangle$, where AF is an altLexConnectiveStructure.

where arg_1, arg_2 are each an **argumentEntity Structure**; other structures defined as follows:

- An **explicitConnectiveStructure** is a pair $\langle$*AttrF, HEAD*$\rangle$, where HEAD is a pair $\langle$*TEXT, CLASS*$\rangle$;
- An **altLexConnectiveStructure** is a pair $\langle$*AttrF, CLASS*$\rangle$;
- an **implicitConnectiveStructure** is a pair $\langle$*AttrF, IC*$\rangle$ or a triple $\langle AttrF, IC_1, IC_2 \rangle$, where IC is a pair $\langle$*TEXT, CLASS*$\rangle$;
- An **argumentEntityStructure** is a triple $\langle$*a, AttrF,SUP*$\rangle$ where a $\in$ ARGS.

In all of the above, AttrF = (ATTR, F) where ATTR is a text span or empty, as defined above; and F = {*s,a,n,i*} with $s \in$ Source, $a \in$ Type, $n \in$ Polarity, $i \in$ Indeterminacy.

4 Concrete syntax

Based on the abstract syntax, a concrete syntax can be defined that provides a physical representation of a PDTB annotation. To ensure meaning-preserving mappings, there should be a one-to-one correspondence between the structures of the concrete syntax and those defined by the corresponding abstract syntax. Correspondingly, the concrete syntax should be mappable to a meaning-preserving representation in GrAF.

Figure 1 shows a concrete XML syntax for the PDTB annotation. Note that explicit and altLex entities, arguments, and supplemental information are anchored in the text by direct references to positions in the primary data (*WSJ* raw text files) and also by Gorn addresses that refer to node(sets) in Penn Treebank constituency trees. Implicit relations, entity relations, and noRels are associated with an *inferenceSite*, which give the character offset of the first character of arg_2 and its sentence number in the primary data.

This structure maps trivially to a GrAF rendering, given in Figure 7. The resulting graph is depicted in Figure 2. Note that the GrAF rendering requires that direct references to primary data in the XML annotation refer instead to *regions* defined in a *base segmentation* document. Such a

```
<explicitRelation
   span="258..262"
   gorn="1,0,1,2,0"
   src="Ot">
  <head sClass="Temp.Asynch.Succ">
      once</head>
  <arg1 span="202..257" gorn="1,0,0;
       1,0,1,0;1,0,1,1;1,0,1,3"
       src="Inh"/>
  <arg2 span="263..282"
        gorn="1,0,1,2,1" src="Inh"/>
</explicitRelation>
```

Figure 1: PDTB Concrete XML syntax

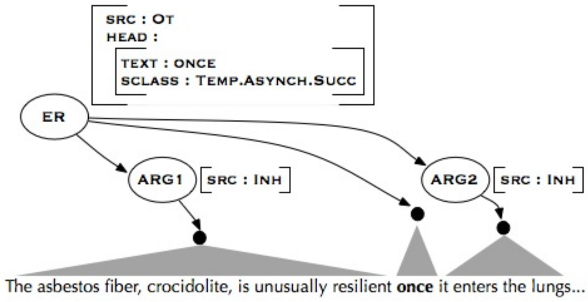

Figure 2: Graphic rendering of PDTB annotation

document would contain the minimal set of regions used by annotations over the data. So, for example, a GrAF rendering of the PTB syntactic annotations, PropBank semantic role annotations,[3] and PDTB annotations over the *WSJ* would refer to the same set of regions, combining several if necessary to refer to less granular or discontiguous spans. This avoids problems of alternative tokenizations, which in turn facilitates the combination of the different layers.

Figure 3 shows the concrete syntax realization of an implicit discourse relation for the text :

6. A Lorillard spokeswoman said, "This is an old story. *We're talking about years ago before anyone heard of asbestos having any questionable properties.* Implicit=besides **There is no asbestos in our products now**." (Expansion.Conjunction/Comparison)

Figure 4 shows the concrete syntax realization of an entity relation for the text :

7. *We have no useful information on whether users are at risk," said James A. Talcott of Boston's Dana-Farber Cancer Institute.* EntRel **Dr. Talcott led a team of researchers from the National Cancer Institute and the medical schools of Harvard University and Boston University**.

[3]The ANC Project has developed PTB-to-GrAF and PropBank-to-GrAF transducers.

```
<implicitRelation
   src="Ot" type="Comm">
   <attr span="726..753"
     gorn="4,0;4,1,0;4,1,1;4,1,2;4,2"/>
  <ic sClass1="Expansion.Conjunction"
      sClass2="Comparison">
      besides</ic>
  <arg1 span="778..874" gorn="5"
       src="Inh"/>
  <arg2 span="876..916"
        gorn="6" src="Inh"/>
</implicitRelation>
```

Figure 3: Concrete syntax for an implicit discourse relation

```
<entityRelation>
<arg1 span="1046..1169"
        gorn="8"/>
  <arg2 span="1171..1311"
        gorn="9"/>
</entityRelation>
```

Figure 4: Concrete syntax for an entity discourse relation

Figures 5 and 6 show their graphic renderings.

5 Discussion

The exercise of creating an abstract syntax for the PDTB annotation scheme and rendering it in a graphic form shows the structure of the annotations clearly. The concrete syntax is much more readable than the original format, and therefore errors and inconsistencies may be more readily identified. Furthermore, because it is rendered in XML, annotations can be validated against an XML schema (including validation that attribute values are among a list of allowable alternatives).

The abstract syntax also suggests an overall structure for a general-purpose standard for annotating discourse relations, in that it identifies a

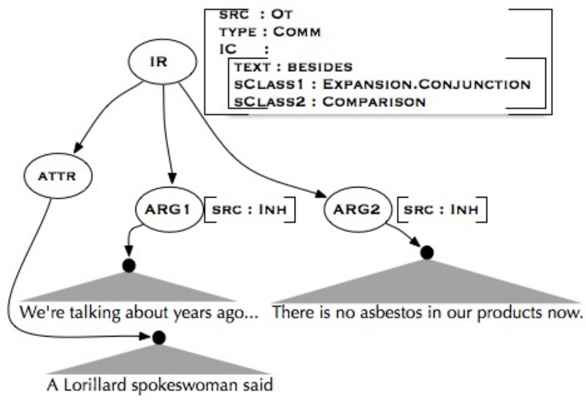

Figure 5: Implicit relation visualization

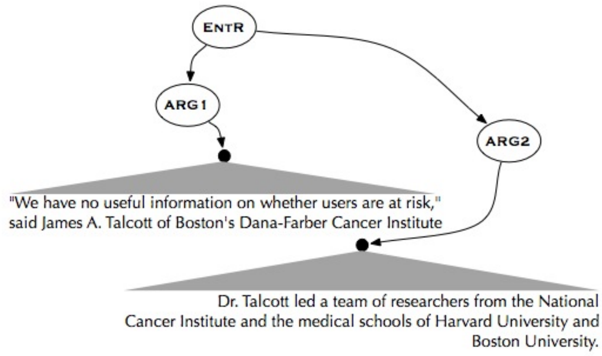

Figure 6: Entity relation visualization

high-level set of relations based on textual realization (the set REL, as given in Section 3) that could provide the top-level of a relation classification scheme. We envision that any general-purpose discourse annotation scheme must allow for annotation based on all or any of several perspectives on elements of the task, such as semantic, interpersonal/intentional, and stylistic/textual identified in Hovy (1995). PDTB annotations are classified as "informational" (semantic, inter-propositional, ideational, pragmatic); the intentional and textual perspectives lie outside the scope of PDTB. PDTB's attribution types and the set of semantic classes[4], combined with those of other schemes, could provide a base for development of a structured set of discourse annotation classes for the ISO specification along the various axes of perspective, and at different levels of granularity.

The work within ISO on discourse relations so far focuses on *discourse graphs* and *discourse trees* that describe discourse structure over an entire text by linking individual relations. Annotating dependencies across relations, however, presumes an understanding of the nature of representation for high-level discourse structure, so that the resulting theory can guide the specification of guidelines for how to compose individual relations. Since there is currently little agreement on such a theory, the PDTB has taken the approach to avoid biasing the annotation towards one or the other theory. Instead, the developers have chosen to specify discourse relations at a low-level, i.e., a level that is clearly defined and well understood. This aspect of the methodology has two related benefits. First, the corpus is usable by researchers across different frameworks for empirical studies on the behavior of low-level discourse relations (for example, studies on the differences between causal relations and contrastive relations). Second, the underlying theory-neutrality itself will allow validation studies of different types of discourse representations (e.g., trees (Polanyi, 1987; Mann and Thompson, 1988; Webber et al., 2003), graphs (Wolf and Gibson, 2005), DAGs (Lee et al., 2008)). In this sense, the PDTB uniquely provides a basis for an *emergent* and *data-driven* theory of discourse structure. Consideration of this approach, either as an alternative to full discourse trees/graphs, or as a base level upon which to build higher-level representations of various types, could be valuable input for the ISO development of a general-purpose standard.

6 Conclusion and Future Work

We have developed an interoperable format for the Penn Discourse Treebank annotation scheme using the strategy for scheme design outlined in Ide and Bunt (2010). We show a high-level XML concrete syntactic realization of the abstract syntax and the corresponding GrAF representation, which ensures that PDTB annotations can be easily combined with other GrAF-based annotation layers in the same source corpus. We are developing a transducer to render the current PDTB annotations in the XML representation, which can be transduced to GrAF using the ANC Tool[5]. This transducer will be made freely available on the PDTB web site.

The underlying annotation framework of the PDTB is followed by several similar discourse annotation projects, in that they are based on a similar discourse annotation framework (e.g., lexicalization of discourse relations, structural independence of relations, and theory neutrality). However, among these schemes there are variations (e.g., the inventory of sense classes, feature sets for attribution, and relation types) dictated by the language and/or domain of the data being annotated. By abstracting out the basic elements of the PDTB scheme and formalizing the structure of the information in a PDTB annotation, we have identified some of the conceptual building blocks of discourse relation analysis that can be used to guide development of new schemes as well as compari-

[4]For a complete list, see the "Penn Discourse Treebank 2.0. Annotation Manual", http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf

[5]http://www.anc.org/tools

```
<node xml:id="pdtb-n101"/>
<a label="explicitRelationEntity"
   ref="pdtb-n101" as="PDTB">
<fs>
  <f name="src" value="Ot"/>
</fs>
</a>

<node xml:id="pdtb-n12"/>
<a label="explicitRelationEntity"
   ref="pdtb-n11" as="PDTB">
<fs>
  <f name="src" value="Ot"/>
  <f name="head">
     <fs>
        <f name="text" value="once"/>
        <f name=sClass"
            value="Temp.Asynch.Succ"/>
     </fs>
   </f>
</fs>
</a>

<edge xml:id="pdtb-e201" from="pdtb-n12"
   to="seg-r15 seg-r16"/>
<edge xml:id="pdtb-e202" from="pdtb-n12"
   to="pdtb-n13"/>
<edge xml:id="pdtb-e203" from="pdtb-n12"
   to="pdtb-n14"/>
<edge xml:id="pdtb-e204" from="pdtb-n12"
   to="pdtb-n15"/>


<node xml:id="pdtb-n14"/>
<a label="arg1" ref="pdtb-n14"
  as="PDTB">
<fs>
  <f name="src" value="Inh"/>
</fs>
</a>

<edge xml:id="pdtb-e205" from="pdtb-n14"
   to="seg-r20"/>

<node xml:id="pdtb-n15"/>
<a label="arg2" ref="pdtb-n15"
  as="PDTB">
<fs>
  <f name="src" value="Inh"/>
</fs>
</a>

<edge xml:id="pdtb-e206" from="pdtb-n15"
   to="seg-r57 seg-r65"/>
```

Figure 7: GrAF rendering of PDTB example

son and combination of schemes. Our immediate goal is to explore the potential of this work to feed the development of an ISO standard for annotating discourse relations.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 17th international conference on Computational linguistics*, pages 86–90, Morristown, NJ, USA. Association for Computational Linguistics.

Harry Bunt. 2010. A methodology for designing semantic annotation languages exploiting semantic-syntactic isomorphisms. In *Proceedings of the Second International Conference on Global Interoperability for Language Resources (ICGL2010)*, pages 29–46, Hong Kong SAR. City University of Hong Kong.

Saul Gorn. 1965. Explicit definitions and linguistic dominoes. In John Hart and Satoru Takasu, editors, *Systems and Computer Science*. University of Toronto Press, Toronto, Canada.

Eduard Hovy. 1995. The multifunctionality of discourse markers. In *Proceedings of the Workshop on Discourse Markers*, The Netherlands. Egmond-aan-Zee.

Nancy Ide and Harry Bunt. 2010. Anatomy of annotation schemes: Mapping to GrAF. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 247–255, Uppsala, Sweden, July. Association for Computational Linguistics.

Nancy Ide and Laurent Romary. 2004. International standard for a linguistic annotation framework. *Journal of Natural Language Engineering*, 10(3–4):211–225.

Nancy Ide and Keith Suderman. 2007. GrAF: A Graph-based Format for Linguistic Annotations. In *Proceedings of the First Linguistic Annotation Workshop*, pages 1–8, Prague.

ISO. 2009. *Semantic annotation framework (SemAF), Part 1: Time and Events. ISO CD 24617-1:2009*. ISO, Geneva.

Alan Lee, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2008. Departures from tree structures in discourse: Shared arguments in the Penn Discourse Treebank. In *Proceedings of the Constraints in Discourse III Workshop*, Potsdam, Germany.

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

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March.

Livia Polanyi. 1987. The linguistic discourse model: Towards a formal theory of discourse structure. Technical Report 6409, Bolt Beranek and Newman, Inc., Cambridge, Mass.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

PDTB-Group. 2008. The Penn Discourse TreeBank 2.0 Annotation Manual. Technical Report IRCS-08-01, Institute for Research in Cognitive Science, University of Pennsylvania.

Bonnie Webber, Aravind Joshi, Matthew Stone, and Alistair Knott. 2003. Anaphora and discourse structure. *Computational Linguistics*, 29(4):545–587.

Florian Wolf and Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2).

FIELD NUM	DESCRIPTION
1.	*Relation Type:* Encodes how the relation is realized in the text: {Explicit, Implicit, AltLex, EntRel, NoRel}
2.	*Conn Span:* Character offsets (or set of offsets for discontinuous text) for the connective or the AltLex text span
3.	*Conn Head:* Head of Explicit connective (provided as text)
4.	*Conn Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Conn Span"
5.	*Conn Attr Src:* Attribution Source features for Connective: {Wr (Writer), Ot (Other), Arb (Arbitrary)}
6.	*Conn Attr Type:* Attribution Type features for Connective: {Comm (assertion), PAtt (belief), Ftv (Fact), Ctrl (action)}
7.	*Conn Attr Pol:* Attribution Polarity feature for Connective: {Neg (negative polarity), Null (no negation)}
8.	*Conn Attr Det:* Attribution Determinacy feature for Connective: {Indet (indeterminate), Null (determinate)}
9.	*Conn Attr Span:* Character offsets (or set of offsets for discontinuous text) for text span of attribution on connective
10.	*Conn Attr Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Conn Feat Span"
11.	*Conn1:* First implicit connective
12.	*SClass1A:* First sense for explicit connective or (first) implicit connective
13.	*SClass1B:* Second sense for explicit connective or (first) implicit connective
14.	*Conn2:* Second implicit connective
15.	*SClass2A:* First sense for second implicit connective
16.	*SClass2B:* Second sense for second implicit connective
17.	*Sup1 Span:* Character offsets (or set of offsets for discontinuous text) for supplemantary text for Arg1
18.	*Sup1 Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Sup1 Span"
19.	*Arg1 Span:* Character offsets (or set of offsets for discontinuous text) for Arg1 text span
20.	*Arg1 Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Arg1 Span"
21.	*Arg1 Attr Src:* Attribution Source features for Arg1: {Wr (Writer), Ot (Other), Arb (Arbitrary), Inh (inherited)}
22.	*Arg1 Attr Type:* Attribution Type features for Arg1: {Comm (assertion), PAtt (belief), Ftv (Fact), Ctrl (actions), Null}
23.	*Arg1 Attr Pol:* Attribution Polarity feature for Arg1: {Neg (negative polarity), Null (no negation)}
24.	*Arg1 Attr Det:* Attribution Determinacy feature for Arg1: {Indet (indeterminate), Null (determinate)}
25.	*Arg1 Attr Span:* Character offsets (or set of offsets) for text span of attribution on Arg1
26.	*Arg1 Attr Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Arg1 Feat Span"
27.	*Arg2 Span:* Character offsets (or set of offsets for discontinuous text) for Arg2 text span
28.	*Arg2 Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Arg2 Span"
29.	*Arg2 Attr Src:* Attribution Source features for Arg2: {Wr (Writer), Ot (Other), Arb (Arbitrary), Inh (inherited)}
30.	*Arg2 Attr Type:* Attribution Type features for Arg2: {Comm (assertion), PAtt (belief), Ftv (Fact), Ctrl (actions), Null}
31.	*Arg2 Attr Pol:* Attribution Polarity feature for Arg2: {Neg (negative polarity), Null (no negation)}
32.	*Arg2 Attr Det:* Attribution Determinacy feature for Arg2: {Indet (indeterminate), Null (determinate)}
33.	*Arg2 Attr Span:* Character offsets (or set of offsets) for text span of attribution on Arg2
34.	*Arg2 Attr Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Arg2 Feat Span"
35.	*Sup2 Span:* Character offsets (or set of offsets for discontinuous text) for supplementary text for Arg2
36.	*Sup2 Gorn:* PTB Gorn address (or set of addresses when text is not covered by a single node) for "Sup2 Span"

Table 1: Annotation Fields for the Penn Discourse Treebank Flat (Horizontal) Format