Enhanced Dialogue Markup for Crisis Talk Scenario Resources

Claudia Sassen & Dafydd Gibbon

Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld
Postfach 10 01 31, D-33501 Bielefeld, Germany
{csassen,gibbon}@spectrum.uni-bielefeld.de

Abstract

We present a method of enhancing dialogue markup by mapping HPSG-based discourse category information into XML. The application scenario is crisis talk, specifically cockpit voice recording (CVR) transcripts of aviation disasters. This approach is new both as a source of richly annotated spoken language corpus resources for a little known scenario, and in grammatical theory and language documentation.

1. Crisis Talk

Crisis talk is defined as a dialogue genre which occurs in threatening situations of unpredictable outcome with no obvious way out, requiring spontaneous decision, unconventional strategies, and unrehearsed actions (Sassen, 2001). Classical spoken language scenarios are typically service encounters and construction dialogues. These are better-behaved genres in the sense of Grice's conversational maxims (Grice, 1975): cooperative, well-formed, and often rehearsed. Unlike these scenarios, crisis talk is more disfluent, violates Gricean maxims, is usually emotional, has high taboo word frequency, uptake loops (Gibbon, 1981), both terminated and unterminated, reprise utterances (Ginzburg et al., 2001), and greater speech output quantity. Crisis talk typically occurs in negotiations with criminals, political summits, interpersonal conflicts, and disaster scenarios.

In aviation communication, crisis talk usually occurs during a plane's take off or landing because during these phases safety is most at risk. The flow of communication may be smooth and professional and thus limited mainly to the speech act types *directive* (order, request) and *assertive* (confirm), as these types feature prominently in the performance of a check list. In most cases, however, the communication is disfluent, and frequent use of *expressives* adds to the inventory of conventional speech act types.

Disfluencies in crisis talk result from speech patterns such as restarts, within-turn repetitions, inter-turn repetitions, and uptake securing mechanisms.

Restarts mark new beginnings in the pronunciation of a word after the speech production was previously broken off, usually resulting in truncated words and/or utterances. The speaker may take up the truncated sign or utterance again and iterate it or he returns to an earlier thought with an extended formulation.

Within-turn repetitions apply to the iteration of language phenomena: on the word level this goes for discourse particles and on the utterance/sentence level for the speech act types *directive* or *assertive*. Especially iterations of directives might be interpreted as a functional shift by which the directive obtains some qualities of an *expressive*. The same might be claimed for iterated discourse particles. This observation goes back to Searle, who makes the critical point that the use of some linguistic expression may have a semantic and functional shift when used in dialogue. This is for instance the case with the English directive verb *urge* which according to Searle has an assertive use, but is in the first instance a directive and as such "to urge is simply to advocate a course of action" (Searle and Vanderveken, 1985), p.200.

Uptake mechanisms help secure understanding between the participants. A special type of uptake securing process is constituted by *uptake loops*, which consist of verbal feedback between interlocutors, i.e. by a disconfirm that marks negative uptake, by a request by which the hearer asks for repetition, clarification or substantiation of the utterance containing the trouble source to which the other reacts with a reply. The loop may be followed by a superordinate positive *uptake*, a confirm, by which the speaker signals understanding. According to Gibbon (1981), the canonical form of an uptake-securing process can be rephrased with *"WHAT? - THAT! - OH..."* Loops may be repeated as often as is necessary for securing understanding (see also Stenström (1994)), whereby an extensive use of loops would mark a breakdown of communication on the object level and indicate a smooth flow on the meta level (see example 5a). It is striking that in many cases of crisis talk the loops are not completed and that after *check* and *clarification* there is no *confirmation*. Instead, the participant who had asked for a repetition talks about something different. The following extracts from CVR transcripts (source: Aviation Safety Network (2000)) instantiate the linguistic phenomena which cause disfluencies. An arrow marks the relevant passage.

An example of restart after a truncated word/utterance and probably return to an earlier thought is:

```
(1) CVR transcript Japan Air Lines Flight 123 -
12 AUG 1985
18:31:35  FE: What? more aft...ah...What was   <-
damaged?
```

Restart after truncated utterance, new thought and probably return to previously truncated utterance:

```
(2) CVR transcript Air Canada Flight 797 -
02 JUN 1983 (19)
CFO120: 19:04:07 Okay I eh, you don't          <-
have to do it now, I can't go back now,
it's too heavy, I think we'd better go down
```

Restart with iteration of last sign of truncated utterance:

```
(3) CVR transcript Lauda Air Flight 004 -
26 MAY 1991 (34)
23.25:26 CA:  Ah, you can tell 'em it,
just it's, it's,  it's, just ah, no, ah,       <-
it's probably ah wa... ah moisture or
something 'cause it's not just, oh, it's
coming on and off.
```

Uptake loops:

```
(4) CVR transcript Lufthansa Flight 2904 -
14 SEP 1993 (44)
15.34:11  PNF    Dreh'n weg (turn it away)      <-
15.34:12  PF     Was? (What?)                   <-
          PNF    Dreh ihn weg (turn it away)    <-
15.34:16  PF     Scheisse! (shit!)
```

```
(5) CVR transcript Avianca Flight 052 -
25 JAN 1990 (31)
CCP126: 2117:40 One zero.
RXX127: 2117:41 Okay, one zero knots, increasing
TWR128: 2117:42 Increase, increase!            <-
CCP129: 2117:42 What?                          <-
RXX130: 2117:44 Increasing                     <-
CCP131: 2117:45 What?                          <-
TWR132: 2117:46 Okay
```

```
(5a) CVR transcript Aviance Flight 052 -
25 JAN 1990 (31)
CCP086: Did you ever take it out of there?     <-
CFO087: Huh?                                   <-
CCP088: Have you ever taken it out of there?   <-
CFO089: Hadn't till now                        <-
```

Within-turn repetition of directives:

```
(6) AeroPeru B757 off Lima (Peru)
2 Oct, 1996
CCO257: 00:52:43 (12:26) THE LOWER             <-
ONE, THE LOWER ONE, THE LOWER ONE, THE         <-
LOWER ONE, THAT LAST ONE... AIR DATA           <-
THERE IT IS.
(...)
CCP260: 00:52:52 (12:35) FUCK..!
BASIC INSTRUMENTS, LET'S GO TO BASIC           <-
INSTRUMENTS!                                   <-
(...)
ATC537: 01:11:02 (30:45) GO UP, GO UP          <-
IF IT INDICATES PULL UP                        <-
CCP538: 01:11:05 (30:48) I HAVE IT, I
HAVE IT!
```

Within-turn repetition of assertives:

```
(7) Birgen Air B757 Accident 6 Feb, 1996
HCP094: 0346:52  thrust levers, thrust
thrust thrust thrust
HCO095: 0346:54  retard
HCP096: 0346:54  thrust, don't pull back,
don't pull back,  don't pullback, don't
pull back
HCO097: 0346:56  okay open open                <-
HCP098: 0346:57  don't pull back,
please don't pull back
HCO099: 0346:59  open sir, open                <-
```

Within-turn repetition of expressives:

```
(8) CVR transcript United Flight 585 - 03 MAR 1991
(...)
CAM027: 09:43:37:4 [Click sound similiar
to that of a flap lever actuation]
CFO028: 09:43:38:4 Oh my God...                <-
[unidentifiable click sound] Oh my God!        <-
```

Inter-turn repetition of expressives:

```
(9) Birgen Air B757 Accident 6 Feb, 1996
(...)
HCP102: 0347:03  what's happening              <-
HCO103: 0347:05  oh what's happening           <-
```

Repetition of discourse particles with the possible function of an expressive:

```
(10) CVR transcript Japan Air Lines Flight 123 -
12 AUG 1985 (23)
18:31:35  FE: What? more aft...ah...What was
damaged?  Where? ah...ah...ah... ah... Coat     <-
room?
```

Other expressives:

```
(11) CVR transcript Lufthansa Flight 2904 -
14 SEP 1993 (44)
          PNF    Dreh ihn weg (turn it away)
15.34:16  PF     Scheisse! (shit!)             <-
```

2. Markup Requirements

For the creation and annotation of a corpus a markup system is necessary. Ideally, it is platform-independent and flexible in that it allows the inclusion of additional elements for special applications. The TEI spoken dialogue transcription guidelines (Sperberg-McQueen and Burnard, 1994) promise to be a suitable starting point. They define a set of generic guidelines for the representation of textual materials in electronic form (poems, drama, spoken dialogue), whether as constituents of a research database or components of non-paper publications. Since the TEI supports loss-free, system-independent interchange, the TEI-scheme would meet the requirement of independence. However, crisis talk scenario requirements are more demanding, and the TEI descriptive elements (Chapter 11) for spoken dialogue transcription are inadequate for crisis talk annotation. A crisis talk markup system clearly needs considerable flexibility. Specifically, crisis talk annotation requires detailed syntactic, semantic and pragmatic features (including deixis, anaphora, speech-act identification, disambiguation and sequencing, and theme-rheme relations). The TEI option, ad hoc DTD modification, is suboptimal, as it is not easy to handle. VERBMOBIL annotation conventions (Gibbon et al. 2000) allow annotation of dialogue acts and some semantic-pragmatic properties, but only cover a small range of the required features, and the notation is not standardised. A principled approach is preferred.

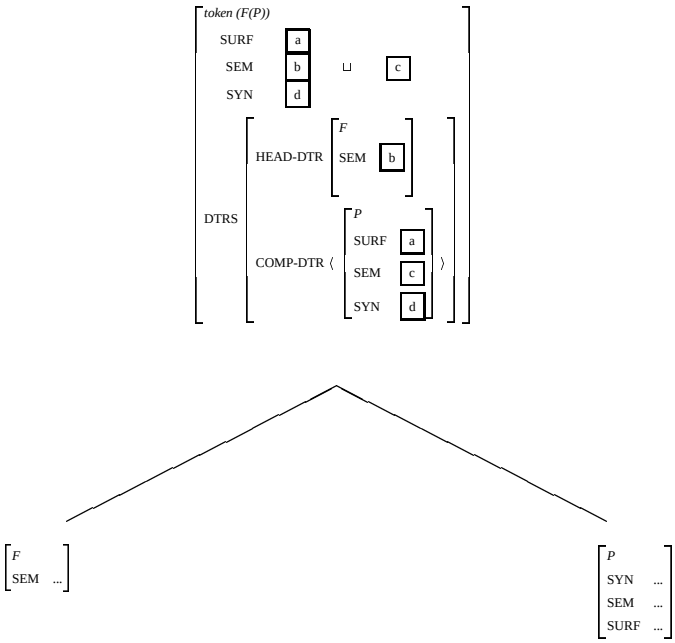

Figure 1: An underspecified tree structure of an illocutionary act that has the *illocutionary force* (*F*) as head-daughter and *proposition* (*P*) as the head-daughter's sister.

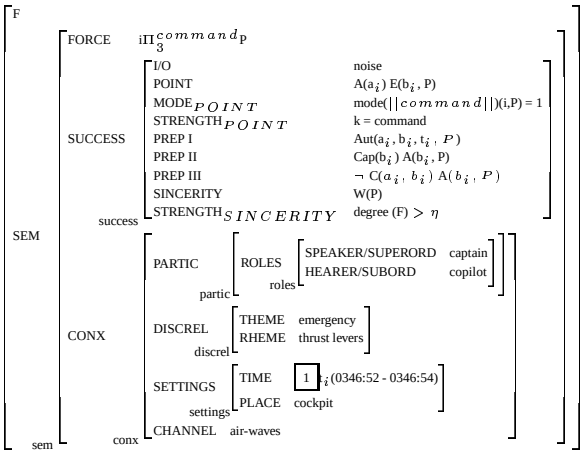

Figure 2: An HPSG-based entry of an item of type *illocutionary force* (*F*). Legend: SUCCESS=conditions of success (Searle and Vanderveken, 1985), I/O=input-output condition, POINT=illocutionary point condition, A=action, a_i=speaker, E=elicitation, b_i=hearer, P=proposition, MODE$_{POINT}$=mode of achievement of the illocutionary point, i=context, STRENGTH$_{POINT}$=degree of strength of the illocutionary point, PREPI–III=preparatory condition, Aut=authority, Cap=capability, C=common knowledge, W=want, STRENGTH$_{SINCERITY}$=degree of strength of the sincerity condition, CONX=context,

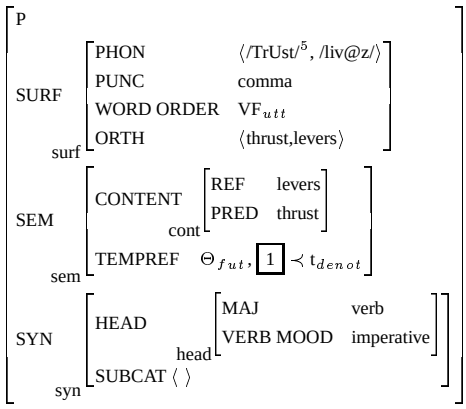

Figure 3: An HPSG-based entry for the item of type *proposition* (*P*). Legend: MAJ=major feature (Pollard and Sag, 1987), REF=reference, PRED=predication, TEMPREF=temporal reference, PHON=phonology, PUNC=punctuation, VF_{utt}=verb of utterance in first position, ORTH=orthography

3. Linguistic Basis for Annotation

One strategy we judge suitable for developing markup extensions in a principled way starts from a linguistic analysis and an attribute-value HPSG-formalism (Pollard and Sag, 1987). However, the rules and principles of HPSG do not capture the structure of many aspects of sentence performance, let alone dialogues, i.e. linguistic structures constituted by more than one speaker. Moreover, natural language dialogues can hardly be captured by HPSG, since it is oriented towards the ideal speaker/hearer of the Chomsky paradigm and not tuned to the performance of imperfect beings. Consequently, the HPSG formalism has to be extended.

An early attempt at modelling spoken language is an HPSG-based model of *Prosody Particle-pairs* that integrates the description of discourse particles and their intonation into a lexicon for spoken language applications (Gibbon and Sassen, 1997). In the resulting HPSG-based sign which is a composite lexical entry the items *prosody* and *discourse particle* jointly function as a complex discourse control sign which has a four-dimensional structure ⟨SYN, DTRS, SURF, SEM⟩ with two *compositional* and two *interpretative* dimensions. The compositional dimensions refer to the syntactic features of the sign such as its distribution in the immediate linguistic context (SYN) and to its internal components of which it is constituted (DTRS). The interpretative dimensions stand for its surface representation (SURF) including aspects of orthography and word order (also its phonetic and perhaps gestural realisation) and for its semantic (SEM) features that include contextual properties.

Figures 1, 2 and 3 jointly endeavour to model the utterance *thrust levers* (see Example 7) from the crisis talk scenario of the Birgen Air B757 Accident (6 Feb, 1996) by an HPSG-based sign. Figure 1, an underspecified structure, takes up the idea of a four-dimensional approach and further extends the HPSG-conventions: it is a complex sign that breaks down into the item of type *illocutionary force (F)* and of type *proposition (P)* whose semantic attributes together form the semantic attribute of the complete sign

(see also Sassen (2001)). While P (see Figure 3) represents the sentence and thus applies to the traditional HPSG-inventory, F (see Figure 2) pertains to the utterance level and surpasses traditional HPSG. New is, for instance, the SUCCESS-attribute. It is included under the SEM-attribute of the illocutionary force and comprises as its substructures Searle's *conditions of success* which serve to unequivocally identify the current speech act (Searle and Vanderveken, 1985). Further essential attributes which have been added are the CONX-attribute, in order to characterise contextual features such as the participants' roles, the discourse relations *theme* and *rheme*, settings and the communication channel. The proposition contains the additional semantic substructures CONT for the content features *reference* and *predication* and TEMPREF in order to determine temporal relations of utterance and action.

The token *thrust levers* has been chosen as the sample utterance since it is a central feature of crisis talk, i.e. part of a turn in which a directive is repeated. As the speaker continues he takes up his directive again in a reduced form in that he iterates the indirect imperative *thrust*, which we categorise for the present purposes as a verb. The repetition is represented by the PHON-attribute as a substructure of SURF whereby the verb is marked by a superscript as a component that occurs more than once. Like uptake loops a repeated directive can be represented as part of a finite state dialogue model.

4. XML as a Denotational Semantics for a Linguistic Description: Application and Results

XML uses an attribute value archiving and retrieval formalism, and is potentially flexible enough to be suitable for fulfilling crisis talk annotation requirements. We define XML attribute-value structures formally as a denotational semantics for an HPSG-type attribute-value description. XML annotation has been criticised for lacking a valid semantics. We handle this problem by using XML simply as algebra for domain structuring in a semantic document model. Together with the appropriate processing mechanisms, XML also provides an operational semantics for the attribute-value description. We have applied this approach to an extensive crisis talk corpus of CVR data that comprises 80 dialogues. Initially, categories were developed heuristically during actual annotation and later formulated in HPSG-style constraints. Starting with a basic XML data annotation, and based on the attribute-value description, an extended DTD was developed and the basic dialogue annotations enhanced semi-automatically. In a sense, our procedure extends, formalises and operationalises the older TEI proposal to formulate markup in terms of feature structures. The following section exemplifies in what way mapping HPSG into XML is achieved. An overview of the steps applied are given first:

- basic XML annotation of a transcribed dialogue fragment,
- creation of the corresponding DTD,
- description of an utterance of the dialogue fragment in HPSG-based, notation (see Figures 1, 2 and 3)
- rewriting the HPSG-based sign as an XML feature structure including its DTD
- enhancing the XML document instance and its DTD
- validation of the DTDs.

A basic XML dialogue annotation of the transcript looks like this:

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE transcript SYSTEM "transcript.dtd">
<transcript>
  <title>Birgen Air B757 Accident 6 Feb, 1996
  </title>
  <speaker>HCP</speaker>
  <turn-id>094</turn-id>
  <time>0346:52</time>
  <turn>thrust levers, thrust thrust thrust thrust
  </turn>
  <speaker>HCO</speaker>
  <turn-id>095</turn-id>
  <time>0346:54</time>
  <turn>retard</turn>
</transcript>
```

The corresponding DTD is:

```
<!ELEMENT transcript (title,(speaker,turn-id,time,turn)+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT speaker (#PCDATA)>
<!ELEMENT turn-id (#PCDATA)>
<!ELEMENT time (#PCDATA)>
<!ELEMENT turn (#PCDATA)>
```

An enriched DTD containing the enhancements of Figures 1, 2 and 3 is formulated as follows:

```
<!ELEMENT transcript (title,(speaker,turn-id,turn)+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT speaker (#PCDATA)>
<!ELEMENT turn-id (#PCDATA)>
<!ELEMENT turn (token)>
<!ELEMENT token (f,p)>
<!ATTLIST token surf IDREF #REQUIRED
                sem  IDREF #REQUIRED
                syn  IDREF #REQUIRED>


 <!ELEMENT f (fsem)>
   <!ELEMENT fsem (force,success,conx)>
  <!ATTLIST fsem occ IDREF #REQUIRED>
   <!ELEMENT force (#PCDATA)>
   <!ELEMENT success (i-o,point,mode-point,strength-point,
                      prepI,prepII,prepIII,sincerity,
                      strength-sincerity)>
    <!ELEMENT i-o (#PCDATA)>
    <!ELEMENT point (#PCDATA)>
    <!ELEMENT mode-point (#PCDATA)>
    <!ELEMENT strength-point (#PCDATA)>
    <!ELEMENT prepI (#PCDATA)>
    <!ELEMENT prepII (#PCDATA)>
    <!ELEMENT prepIII (#PCDATA)>
    <!ELEMENT sincerity  (#PCDATA)>
    <!ELEMENT strength-sincerity (#PCDATA)>
    <!ELEMENT conx (partic,discrel,settings,channel)>
  <!ELEMENT partic (roles)>
   <!ELEMENT roles (speakersubord,hearersubord)>
    <!ELEMENT speakersubord (#PCDATA)>
    <!ELEMENT hearersubord (#PCDATA)>
  <!ELEMENT discrel (theme,rheme)>
   <!ELEMENT theme (#PCDATA)>
   <!ELEMENT rheme (#PCDATA)>
  <!ELEMENT settings (time?,place)>
   <!ELEMENT time (#PCDATA)>
   <!ATTLIST time occ ID #REQUIRED>
   <!ELEMENT place (#PCDATA)>
  <!ELEMENT channel (#PCDATA)>

 <!ELEMENT p (psurf,psem,psyn)>
   <!ELEMENT psurf (phon,punc?,word-order,orth)>
```

```
  <!ATTLIST psurf occ ID #REQUIRED>
   <!ELEMENT phon (#PCDATA)>
   <!ELEMENT punc (#PCDATA)>
   <!ELEMENT word-order (#PCDATA)>
   <!ELEMENT orth (#PCDATA)>
  <!ELEMENT psem (content,tempref)>
  <!ATTLIST psem occ ID #REQUIRED>
    <!ELEMENT content (ref,pred)>
     <!ELEMENT ref (#PCDATA)>
     <!ELEMENT pred (#PCDATA)>
    <!ELEMENT tempref (#PCDATA)>
    <!ATTLIST tempref occ IDREF #IMPLIED>
  <!ELEMENT psyn (head,subcat)>
  <!ATTLIST psyn occ ID #REQUIRED>
    <!ELEMENT head (maj,verb-mood)>
     <!ELEMENT maj (#PCDATA)>
     <!ELEMENT verb-mood (#PCDATA)>
    <!ELEMENT subcat (#PCDATA)>
```

Note that with the enhancement some modifications were necessary in order to maintain a correct syntax: the *time* element of the simple DTD has been integrated in the substructures of the *settings* element, so that it does not appear any more in the content model of the *transcript* element. Here is a fragment of the transcript with enhanced markup:

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE transcript SYSTEM "transcript.dtd">
<transcript>
  <title>Birgen Air B757 Accident 6 Feb, 1996
  </title>
  <speaker>HCP</speaker>
  <turn-id>094</turn-id>
   <turn>
    <token surf="d" sem="a" syn="c">
 <f>
    </fsurf-->
 <fsem occ="a">
   <force>i Pi-3^command P</force>
   <success>
     <i-o> noise </i-o>
     <point>A(a-i) E(b-i,P)=1</point>
     <mode-point> mode(||command||)(i,P)=1</mode-point>
     <strength-point> k=command   </strength-point>
     <prepI> Aut(a-i,b-i,t-i,P)</prepI>
     <prepII> Cap(b-i)A(b-i,P)</prepII>
     <prepIII>not C(a-i,b-i)A(b-i,P)</prepIII>
     <sincerity>W(P) </sincerity>
     <strength-sincerity>degree(F)>''n''
     </strength-sincerity>
   </success>
   <conx>
     <partic>
     <roles>
       <speakersubord>captain</speakersubord>
       <hearersubord>copilot</hearersubord>
     </roles>
     </partic>
     <discrel>
       <theme> emergency</theme>
       <rheme>thrust levers</rheme>
     </discrel>
     <settings>
       <time occ="b">(0346:52-0346:54)</time>
       <place>cockpit</place>
     </settings>
     <channel>air-waves</channel>
   </conx>
  </fsem>
 </f>
 <p>
  <psurf occ="d">
    <phon>/diskonekt/, /Di/, /Ot@UpaIl@t/</phon>
    <punc> comma</punc>
    <word-order>VF-utt</word-order>
    <orth>thrust, levers</orth>
  </psurf>
  <psem occ="a">
    <content>
      <ref>levers</ref>
      <pred>thrust</pred>
    </content>
    <tempref occ="b"> delta-tfut prec t-denot</tempref>
  </psem>
  <psyn occ="c">
   <head>
    <maj>verb </maj>
    <verb-mood> imperative </verb-mood>
   </head>
    <subcat> empty </subcat>
  </psyn>
 </p>
</token>
  </turn>
  <speaker>HCO</speaker>
  <turn-id>095</turn-id>
  <time>0346:54</time>
  <turn>
   <token><...>...</...>retard
   </token>
  </turn>
</transcript>
```

The DTDs have been operationally validated with an nsgmls parser.

5. Conclusion

Using an extensive crisis talk corpus we introduce a principled and flexible strategy for developing a new annotation category set by mapping HPSG-based attribute value matrices into an XML semantics. We show that this strategy has the power and flexibility to handle a complex new and highly unorthodox scenario of unrehearsed and dangerous real-life situations. Current work in this area is directed towards validating and extending the markup categories with additional data sets.

6. References

Dafydd Gibbon and Claudia Sassen. 1997. Prosody particle-pairs as discourse control signs. *EUROSPEECH Proceedings*, I.

Dafydd Gibbon. 1981. Idiomaticity and functional variation. A case study of international amateur radio talk. *Language in Society*, 10:21–42.

Jonathan Ginzburg, Ivan Sag, and Matthew Purver. 2001. Integrating conversational move types in the grammar of conversation. *BI-DIALOG Proceedings*, pages 45–56.

Henry Paul Grice. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntacs and Semantics*, volume 3 Speech Acts. Academic Press, New York.

Aviation Safety Network. 2000. CVR and ATC transcripts. http://aviation-safety.net/cvr/transcripts.htm.

Carl Jesse Pollard and Ivan Sag. 1987. *Information-based syntax and semantics, Vol.I: Fundamentals*. Stanford: Center for the Study of Language and Information.

Claudia Sassen. 2001. An HPSG-based representation model for illocutionary acts in crisis talk. *BI-DIALOG Proceedings*, pages 219–227.

John Searle and Daniel Vanderveken. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Michael Sperberg-McQueen and Lou Burnard, editors. 1994. *Guidelines for electronic text encoding and interchange (TEI P3)*. Chicago and Oxford: Association for Computers and the Humanities/Association for Computational Linguistics/Association for Literary and Linguisitc Computing.

Anna-Brita Stenström. 1994. *An introduction to spoken interaction*. Longman.