What lies beneath: Semantic and syntactic analysis of manually reconstructed spontaneous speech

Erin Fitzgerald
Johns Hopkins University
Baltimore, MD, USA
erinf@jhu.edu

Frederick Jelinek
Johns Hopkins University
Baltimore, MD, USA
jelinek@jhu.edu

Robert Frank
Yale University
New Haven, CT, USA
bob.frank@yale.edu

Abstract

Spontaneously produced speech text often includes disfluencies which make it difficult to analyze underlying structure. Successful reconstruction of this text would transform these errorful utterances into fluent strings and offer an alternate mechanism for analysis.

Our investigation of naturally-occurring spontaneous speaker errors aligned to corrected text with manual semantico-syntactic analysis yields new insight into the syntactic and structural semantic differences between spoken and reconstructed language.

1 Introduction

In recent years, natural language processing tasks such as machine translation, information extraction, and question answering have been steadily improving, but relatively little of these systems besides transcription have been applied to the most natural form of language input: spontaneous speech. Moreover, there has historically been little consideration of how to analyze the underlying semantico-syntactic structure of speech.

A system would accomplish *reconstruction* of its spontaneous speech input if its output were to represent, in flawless, fluent, and content-preserved English, the message that the speaker intended to convey (Fitzgerald and Jelinek, 2008; Fitzgerald et al., 2009). Examples of such reconstructions are seen in the following sentence-like units (SUs).

EX1: that's uh that's a relief
becomes
that's a relief

EX2: how can you do that without + it's a catch-22
becomes
how can you do that without <ARG>
it's a catch-22

EX3: they like video games some kids do
becomes
some kids like video games

In EX1, reconstruction requires only the deletion of a simple filled pause and speaker repetition (or *reparandum* (Shriberg, 1994)). The second example shows a *restart fragment*, where an utterance is aborted by the speaker and then restarted with a new train of thought. Reconstruction here requires

1. Detection of an *interruption point* (denoted + in the example) between the abandoned thought and its replacement,
2. Determination that the abandoned portion contains unique and preservable content and should be made a new sentence rather than be deleted (which would alter meaning)
3. Analysis showing that a required argument must be inserted in order to complete the sentence.

Finally, in the third example EX3, in order to produce one of the reconstructions given, a system must

1. Detect the anaphoric relationship between "they" and "some kids"
2. Detect the referral of "do" to "like video games"
3. Make the necessary word reorderings and deletion of the less informative lexemes.

These examples show varying degrees of difficulty for the task of automatic reconstruction. In each case, we also see that semantic analysis of the reconstruction is more straightforward than of the original string directly. Such analysis not only informs us of what the speaker intended to communicate, but also reveals insights into the types of errors speakers make when speaking spontaneously and where these errors occur. The semantic labeling of reconstructed sentences, when combined with the reconstruction alignments, may yield new quantifiable insights into the structure of disfluent natural speech text.

In this paper, we will investigate this relationship further. Generally, we seek to answer two questions:

- What generalizations about the underlying structure of errorful and reconstructed speech utterances are possible?
- Are these generalizations sufficiently robust as to be incorporated into statistical models in automatic systems?

We begin by reviewing previous work in the automatic labeling of structural semantics and motivating the analysis not only in terms of discovery but also regarding its potential application to automatic speech reconstruction research. In Section 2 we describe the Spontaneous Speech Reconstruction (SSR) corpus and the manual semantic role labeling it includes. Section 3 analyzes structural differences between verbatim and reconstructed text in the SSR as evaluated by a combination of manual and automatically generated phrasal constituent parses, while Section 4 combines syntactic structure and semantic label annotations to determine the consistency of patterns and their comparison to similar patterns in the Wall Street Journal (WSJ)-based Proposition Bank (PropBank) corpus (Palmer et al., 2005). We conclude by offering a high level analysis of discoveries made and suggesting areas for continued analysis in the future. Expanded analysis of these results is described in (Fitzgerald, 2009).

1.1 Semantic role labeling

Every verb can be associated with a set of core and optional argument roles, sometimes called a **roleset**. For example, the verb "say" must have a *sayer* and an *utterance which is said*, along with an optionally defined *hearer* and any number of locative, temporal, manner, etc. adjunctival arguments.

The task of predicate-argument labeling (sometimes called semantic role labeling or SRL) assigns a simple *who* did *what* to *whom when*, *where*, *why*, *how*, etc. structure to sentences (see Figure 1), often for downstream processes such as information extraction and question answering. Reliably identifying and assigning these roles to grammatical text is an active area of research (Gildea and Jurafsky, 2002; Pradhan et al., 2004; Pradhan et al., 2008), using training resources like the Linguistic Data Consortium's Proposition Bank (PropBank) (Palmer et al., 2005), a 300k-word corpus with semantic role relations labeled for verbs in the WSJsection of the Penn Treebank.

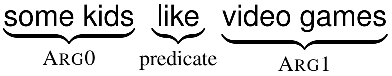

Figure 1: Semantic role labeling for the sentence "some kids like video games". According to PropBank specifications, core arguments for each predicate are assigned a corresponding label ARG0-ARG5 (where ARG0 is a proto-agent, ARG1 is a proto-patient, etc. (Palmer et al., 2005)).

A common approach for automatic semantic role labeling is to separate the process into two steps: argument identification and argument labeling. For each task, standard cue features in automatic systems include verb identification, analysis of the syntactic path between that verb and the prospective argument, and the direction (to the left or to the right) in which the candidate argument falls in respect to its predicate. In (Gildea and Palmer, 2002), the effect of parser accuracy on semantic role labeling is quantified, and consistent quality parses were found to be essential when automatically identifying semantic roles on WSJ text.

1.2 Potential benefit of semantic analysis to speech reconstruction

With an adequate amount of appropriately annotated conversational text, methods such as those referred to in Section 1.1 may be adapted for transcriptions of spontaneous speech in future research. Furthermore, given a set of semantic role labels on an ungrammatical string, and armed with the knowledge of a set of core semantico-syntactic principles which constrain the set of possible grammatical sentences, we hope to discover and take advantage of new cues for construction errors in the field of automatic spontaneous speech reconstruction.

2 Data

We conducted our experiments on the Spontaneous Speech Reconstruction (SSR) corpus (Fitzgerald and Jelinek, 2008), a 6,000 SU set of reconstruction annotations atop a subset of Fisher conversational telephone speech data (Cieri et al., 2004), including

- manual word alignments between corresponding original and cleaned sentence-like units (SUs) which are labeled with transformation types (Section 2.1), and
- annotated semantic role labels on predicates and their arguments for all grammatical reconstructions (Section 2.2).

The fully reconstructed portion of the SSR corpus consists of 6,116 SUs and 82,000 words total. While far smaller than the 300,000-word PropBank corpus, we believe that this data will be adequate for an initial investigation to characterize semantic structure of verbatim and reconstructed speech.

2.1 Alignments and alteration labels

In the SSR corpus, words in each reconstructed utterance were deleted, inserted, substituted, or moved as required to make the SU as grammatical as possible without altering the original meaning and without the benefit of extrasentential context. Alignments between the original words and their reconstructed "source" words (i.e. in the noisy channel paradigm) are explicitly defined, and for each alteration a corresponding *alteration label* has been chosen from the following.

- DELETE words: fillers, repetitions/revisions, false starts, co-reference, leading conjugation, and extraneous phrases
- INSERT neutral elements, such as *function words* like "the", *auxiliary verbs* like "is", or undefined *argument placeholders*, as in "he wants <ARG>"
- SUBSTITUTE words to *change tense or number*, *correct transcriber errors*, and *replace colloquial phrases* (such as: "he was like..." → "he said...")
- REORDER words (within sentence boundaries) and label as *adjuncts*, *arguments*, or *other structural reorderings*

Unchanged original words are aligned to the corresponding word in the reconstruction with an arc marked BASIC.

2.2 Semantic role labeling in the SSR corpus

One goal of speech reconstruction is to develop machinery to automatically reduce an utterance to its underlying meaning and then generate clean text. To do this, we would like to understand how semantic structure in spontaneous speech text varies from that of written text. Here, we can take advantage of the semantic role labeling included in the SSR annotation effort.

Rather than attempt to label incomplete utterances or errorful phrases, SSR annotators assigned semantic annotation only to those utterances which were well-formed and grammatical post-reconstruction. Therefore, only these utterances (about 72% of the annotated SSR data) can be given a semantic analysis in the following sections. For each well-formed and grammatical sentence, all (non-auxiliary and non-modal) verbs were identified by annotators and the corresponding predicate-argument structure was labeled according to the role-sets defined in the PropBank annotation effort[1].

We believe the transitive bridge between the aligned original and reconstructed sentences and the predicate-argument labels for those reconstructions (described further in Section 4) may yield insight into the structure of speech errors and how to extract these verb-argument relationships in verbatim and errorful speech text.

3 Syntactic variation between original and reconstructed strings

As we begin our analysis, we first aim to understand the types of syntactic changes which occur during the course of spontaneous speech reconstruction. These observations are made empirically given automatic analysis of the SSR corpus annotations. Syntactic evaluation of speech and reconstructed structure is based on the following resources:

1. the manual parse P_{v_m} for each verbatim original SU (from SSR)
2. the automatic parse P_{v_a} of each verbatim original SU

[1] PropBank roleset definitions for given verbs can be reviewed at http://www.cs.rochester.edu/∼gildea/Verbs/.

3. the automatic parse P_{r_a} of each reconstructed SU

We note that automatic parses (using the state of the art (Charniak, 1999) parser) of verbatim, unreconstructed strings are likely to contain many errors due to the inconsistent structure of verbatim spontaneous speech (Harper et al., 2005). While this limits the reliability of syntactic observations, it represents the current state of the art for syntactic analysis of unreconstructed spontaneous speech text.

On the other hand, automatically obtained parses for cleaned reconstructed text are more likely to be accurate given the simplified and more predictable structure of these SUs. This observation is unfortunately not evaluable without first manually parsing all reconstructions in the SSR corpus, but is assumed in the course of the following syntax-dependent analysis.

In reconstructing from errorful and disfluent text to clean text, a system makes not only surface changes but also changes in underlying constituent dependencies and parser interpretation. We can quantify these changes in part by comparing the internal context-free structure between the two sets of parses.

We compare the internal syntactic structure between sets P_{v_a} and P_{r_a} of the SSR check set. Statistics are compiled in Table 1 and analyzed below.

- 64.2% of expansion rules in parses P_{v_a} also occur in reconstruction parses P_{r_a}, and 92.4% (86.8%) of reconstruction parse P_{r_a} expansions come directly from the verbatim parses P_{v_a} (from columns one and two of Table 1).
- Column three of Table 1 shows the rule types most often dropped from the verbatim string parses P_{v_a} in the transformation to reconstruction. The P_{v_a} parses select full clause non-terminals (NTs) for the verbatim parses which are not in turn selected for automatic parses of the reconstruction (e.g. [SBAR → S] or [S → VP]). This suggests that these rules may be used to handle errorful structures not seen by the trained grammar.
- Rule types in column four of Table 1 are the most often "generated" in P_{r_a} (as they are unseen in the automatic parse P_{v_a}). Since rules like [S → NP VP], [PP → IN NP], and [SBAR → IN S] appear in a reconstruction parse but not corresponding verbatim parse at similar frequencies regardless of whether P_{v_m} or P_{v_a} are being compared, we are more confident that these patterns are effects of the verbatim-reconstruction comparison and not the specific parser used in analysis. The fact that these patterns occur indicates that it is these common rules which are most often confounded by spontaneous speaker errors.
- Given a Levenshtein alignment between altered rules, the most common changes within a given NT phrase are detailed in column five of Table 1. We see that the most common aligned rule changes capture the most basic of errors: a leading coordinator (#1 and 2) and rules proceeded by unnecessary filler words (#3 and 5). Complementary rules #7 and 9 (e.g. VP → [*rule*]/[*rule* SBAR] and VP → [*rule* SBAR]/[*rule*]) show that complementing clauses are both added and removed, possibly in the same SU (i.e. a phrase shift), during reconstruction.

4 Analysis of semantics for speech

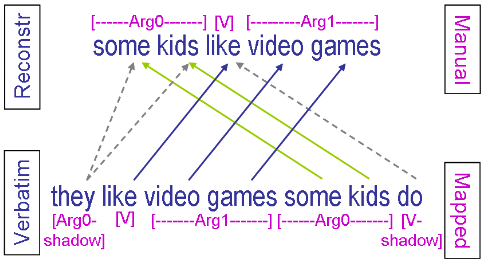
Reconstr
[------Arg0-------] [V] [---------Arg1-------]
some kids like video games
Manual
Verbatim
they like video games some kids do
[Arg0-shadow] [V] [-------Arg1-------] [------Arg0-------] [V-shadow]
Mapped

Figure 2: Manual semantic role labeling for the sentence "some kids like video games" and SRL mapped onto its verbatim source string "they like video games and stuff some kids do"

To analyze the semantic and syntactic patterns found in speech data and its corresponding reconstructions, we project semantic role labels from strings into automatic parses, and moreover from their post-reconstruction source to the verbatim original speech strings via the SSR manual word alignments, as shown in Figures 2.

The automatic SRL mapping procedure from the reconstructed string W_r to related parses P_{r_a} and P_{v_a} and the verbatim original string W_v is as follows.

P_{v_a} rules in P_{r_a}	P_{r_a} rules in P_{v_a}	P_{v_a} rules most frequently dropped	P_{r_a} rules most frequently added	Levenshtein-aligned expansion changes (P_{v_a}/P_{r_a})
64.2%	92.4%	1. NP → PRP	1. S → NP VP	1. S → [CC *rule*] / [*rule*]
		2. ROOT → S	2. PP → IN NP	2. S → [CC NP VP] / [NP VP]
		3. S → NP VP	3. ROOT → S	3. S → [INTJ *rule*] / [*rule*]
		4. INTJ → UH	4. ADVP → RB	4. S → [NP *rule*] / [*rule*]
		5. PP → IN NP	5. S → NP ADVP VP	5. S → [INTJ NP VP] / [NP VP]
		6. ADVP → RB	6. SBAR → IN S	6. S → [NP NP VP] / [NP VP]
		7. SBAR → S	7. SBAR → S	7. VP → [*rule*] / [*rule* SBAR]
		8. NP → DT NN	8. S → ADVP NP VP	8. S → [RB *rule*] / [*rule*]
		9. S → VP	9. S → VP	9. VP → [*rule* SBAR] / [*rule*]
		10. PRN → S	10. NP → NP SBAR	10. S → [*rule*] / [ADVP *rule*]

Table 1: Internal syntactic structure removed and gained during reconstruction. This table compares the rule expansions for each verbatim string automatically parsed P_{v_a} and the automatic parse of the corresponding reconstruction in the SSR corpus (P_{r_a}).

1. **Tag** each reconstruction word $w_r \in$ string W_r with the annotated SRL tag t_{w_r}.
   (a) Tag each verbatim word $w_v \in$ string W_v aligned to w_r via a BASIC, REORDER, or SUBSTITUTE alteration label with the SRL tag t_{w_r} as well.
   (b) Tag each verbatim word w_v aligned to w_r via a DELETE REPETITION or DELETE CO-REFERENCE alignment with a *shadow* of that SRL tag t_{w_r} (see the lower tags in Figure 2 for an example)

   Any verbatim original word w_v with any other alignment label is ignored in this semantic analysis as SRL labels for the aligned reconstruction word w_r do not directly translate to them.
2. **Overlay** tagged words of string W_v and W_r with the automatic (or manual) parse of the same string.
3. **Propagate** labels. For each constituent in the parse, if all children within a syntactic constituent expansion (or all but EDITED or INTJ) has a given SRL tag for a given predicate, we instead tag that NT (and not children) with the semantic label information.

4.1 Labeled verbs and their arguments

In the 3,626 well-formed and grammatical SUs labeled with semantic roles in the SSR, 895 distinct verb types were labeled with core and adjunct arguments as defined in Section 1.1. The most frequent of these verbs was the orthographic form "'s" which was labeled 623 times, or in roughly 5% of analyzed sentences. Other forms of the verb "to be", including "is", "was", "be", "are", "re", "'m", and "being", were labeled over 1,500 times, or at a rate of nearly one in half of all well-formed reconstructed sentences. The verb type frequencies roughly follow a Zipfian distribution (Zipf, 1949), where most verb words appear only once (49.9%) or twice (16.0%).

On average, 1.86 core arguments (ARG[0-4]) are labeled per verb, but the specific argument types and typical argument numbers per predicate are verb-specific. For example, the ditransitive verb "give" has an average of 2.61 core arguments for its 18 occurrences, while the verb "divorced" (whose core arguments "initiator of end of marriage" and "ex-spouse" are often combined, as in "we divorced two years ago") was labeled 11 times with an average of 1.00 core arguments per occurrence.

In the larger PropBank corpus, annotated atop WSJ news text, the most frequently reported verb root is "say", with over ten thousand labeled appearances in various tenses (this is primarily explained by the genre difference between WSJ and telephone speech)[2]; again, most verbs occur two or fewer times.

4.2 Structural semantic statistics in cleaned speech

A reconstruction of a verbatim spoken utterance can be considered an underlying form, analogous

[2] The reported PropBank analysis ignores past and present participle (passive) usage; we do not do this in our analysis.

to that of Chomskian theory or Harris's conception of transformation (Harris, 1957). In this view, the original verbatim string is the surface form of the sentence, and as in linguistic theory should be constrained in some manner similar to constraints between Logical Form (LF) and Surface Structure (SS).

Data	SRL	Total	Most common syntactic categories, with rel. frequency	
P_{v_a}	ARG1	10110	NP (50%)	PP (6%)
P_{r_a}		8341	NP (58%)	SBAR (9%)
PB05			Obj-NP (52%)	S (22%)
P_{v_a}	ARG0	4319	NP (90%)	WHNP (3%)
P_{r_a}		4518	NP (93%)	WHNP (3%)
PB05			Subj-NP (97%)	NP (2%)
P_{v_a}	ARG2	3836	NP (28%)	PP (13%)
P_{r_a}		3179	NP (29%)	PP (18%)
PB05			NP (36%)	Obj-NP (29%)
P_{v_a}	TMP	931	ADVP (25%)	NP (20%)
P_{r_a}		872	ADVP (27%)	PP (18%)
PB05			ADVP (26%)	PP-in (16%)
P_{v_a}	MOD	562	MD (58%)	TO (18%)
P_{r_a}		642	MD (57%)	TO (19%)
PB05			MD (99%)	ADVP (1%)
P_{v_a}	LOC	505	PP (47%)	ADVP (16%)
P_{r_a}		489	PP (54%)	ADVP (17%)
PB05			PP-in (59%)	PP-on (10.0%)

Table 2: Most frequent phrasal categories for common arguments in the SSR (mapping SRLs onto P_{v_a} parses). *PB05* refers to the PropBank data described in (Palmer et al., 2005).

Data	NT	Total	Most common argument labels, with rel. frequency	
P_{v_a}	NP	10541	ARG1 (48%)	ARG0 (37%)
P_{r_a}		10218	ARG1 (47%)	ARG0 (41%)
PB05			ARG2 (34%)	ARG1 (24%)
PB05	Subj-NP		ARG0 (79%)	ARG1 (17%)
PB05	Obj-NP		ARG1 (84%)	ARG2 (10%)
P_{v_a}	PP	1714	ARG1 (34%)	ARG2 (30%)
P_{r_a}		1777	ARG1 (31%)	ARG2 (30%)
PB05	PP-in		LOC (48%)	TMP (35%)
PB05	PP-at		EXT (36%)	LOC (27%)
P_{v_a}	ADVP	1519	ARG2 (21%)	ARG1 (19%)
P_{r_a}		1444	ARG2 (22%)	ADV (20%)
PB05			TMP (30%)	MNR (22%)
P_{v_a}	SBAR	930	ARG1 (61%)	ARG2 (14%)
P_{r_a}		1241	ARG1 (62%)	ARG2 (12%)
PB05			ADV (36%)	TMP (30%)
P_{v_a}	S	523	ARG1 (70%)	ARG2 (16%)
P_{r_a}		526	ARG1 (72%)	ARG2 (17%)
PB05			ARG1 (76%)	ADV (9%)
P_{v_a}	MD	449	MOD (73%)	ARG1 (18%)
P_{r_a}		427	MOD (86%)	ARG1 (11%)
PB05			MOD (97%)	Adjuncts (3%)

Table 3: Most frequent argument categories for common syntactic phrases in the SSR (mapping SRLs onto P_{v_a} parses).

In this section, we identify additional trends which may help us to better understand these constraints, such as the most common phrasal category for common arguments in common contexts – listed in Table 2 – and the most frequent semantic argument type for NTs in the SSR – listed in Table 3.

4.3 Structural semantic differences between verbatim speech and reconstructed speech

We now compare the placement of semantic role labels with reconstruction-type labels assigned in the SSR annotations.

These analyses were conducted on P_{r_a} parses of reconstructed strings, the strings upon which semantic labels were directly assigned.

Reconstructive deletions

Q: Is there a relationship between speaker error types requiring deletions and the argument shadows contained within? Only two deletion types – repetitions/revisions and co-references – have direct alignments between deleted text and preserved text and thus can have argument shadows from the reconstruction marked onto the verbatim text.

Of 9,082 propagated *deleted repetition/ revision* phrase nodes from P_{v_a}, we found that 31.0% of arguments within were ARG1, 22.7% of arguments were ARG0, 8.6% of nodes were predicates labeled with semantic roles of their own, and 8.4% of argument nodes were ARG2. Just 8.4% of "delete repetition/revision" nodes were modifier (vs. core) arguments, with TMP and CAU labels being the most common.

Far fewer (353) nodes from P_{v_a} represented *deleted co-reference* words. Of these, 57.2% of argument nodes were ARG1, 26.6% were ARG0 and 13.9% were ARG2. 7.6% of "argument" nodes here were SRL-labeled predicates, and 10.2% were in modifier rather than core arguments, the most prevalent were TMP and LOC.

These observations indicate to us that redundant co-references are far most likely to occur for ARG1 roles (most often objects, though also subjects for copular verbs (i.e. "to be") and others) and appear more likely than random to occur in core argument regions of an utterance rather than in optional modifying material.

Reconstructive insertions

Q: When null arguments are inserted into reconstructions of errorful speech, what semantic role do they typically fill? Three types of insertions were made by annotators during the reconstruction of the SSR corpus. *Inserted function words*, the most common, were also the most varied. Analyzing the automatic parses of the reconstructions P_{r_a}, we find that the most commonly assigned parts-of-speech (POS) for these elements was fittingly `IN` (21.5%, preposition), `DT` (16.7%, determiner) and `CC` (14.3%, conjunction). Interestingly, we found that the next most common POS assignments were noun labels, which may indicate errors in SSR labeling.

Other inserted word types were auxiliary or otherwise *neutral verbs*, and, as expected, most POS labels assigned by the parses were verb types, mostly `VBP` (non-third person present singular). About half of these were labeled as predicates with corresponding semantic roles; the rest were unlabeled which makes sense as true auxiliary verbs were not labeled in the process.

Finally, around 147 insertion types made were *neutral arguments* (given the orthographic form `<ARG>`). 32.7% were common nouns and 18.4% of these were labeled personal pronouns `PRP`. Another 11.6% were adjectives `JJ`. We found that 22 (40.7%) of 54 neutral argument nodes directly assigned as semantic roles were ARG1, and another 33.3% were ARG0. Nearly a quarter of inserted arguments became part of a larger phrase serving as a modifier argument, the most common of which were CAU and LOC.

Reconstructive substitutions

Q: How often do substitutions occur in the analyzed data, and is there any semantic consistency in the types of words changed? 230 phrase *tense substitutions* occurred in the SSR corpus. Only 13 of these were directly labeled as predicate arguments (as opposed to being part of a larger argument), 8 of which were ARG1. Morphology changes generally affect verb tense rather than subject number, and with no real impact on semantic structure.

Colloquial substitutions of verbs, such as "he was like..." → "he said...", yield more unusual semantic analysis on the unreconstructed side as non-verbs were analyzed as verbs.

Reconstructive word re-orderings

Q: How is the predicate-argument labeling affected? If reorderings occur as a phrase, what type of phrase? Word reorderings labeled as *argument movements* occurred 136 times in the 3,626 semantics-annotated SUs in the SSR corpus. Of these, 81% were directly labeled as arguments to some sentence-internal predicate. 52% of those arguments were ARG1, 17% were ARG0, and 13% were predicates. 11% were labeled as modifying arguments rather than core arguments, which may indicate confusion on the part of the annotators and possibly necessary cleanup.

More commonly labeled than argument movement was *adjunct movement*, assigned to 206 phrases. 54% of these reordered adjuncts were not directly labeled as predicate arguments but were within other labeled arguments. The most commonly labeled adjunct types were TMP (19% of all arguments), ADV (13%), and LOC (11%).

Syntactically, 25% of reordered adjuncts were assigned `ADVP` by the automatic parser, 19% were assigned `NP`, 18% were labeled `PP`, and remaining common NT assignments included `IN`, `RB`, and `SBAR`.

Finally, 239 phrases were labeled as being reordered for the general reason of *fixing the grammar*, the default change assignment given by the annotation tool when a word was moved. This category was meant to encompass all movements not included in the previous two categories (arguments and adjuncts), including moving "I guess" from the middle or end of a sentence to the beginning, determiner movement, etc. Semantically, 63% of nodes were directly labeled as predicates or predicate arguments. 34% of these were PRED, 28% were ARG1, 27% were ARG0, 8% were ARG2, and 8% were roughly evenly distributed across the adjunct argument types.

Syntactically, 31% of these changes were `NP`s, 16% were `ADVP`s, and 14% were `VBP`s (24% were verbs in general). The remaining 30% of changes were divided amongst 19 syntactic categories from `CC` to `DT` to `PP`.

4.4 Testing the generalizations required for automatic SRL for speech

The results described in (Gildea and Palmer, 2002) show that parsing dramatically helps during the course of automatic SRL. We hypothesize that the current state-of-art for parsing speech is adequate to generally identify semantic roles in spontaneously produced speech text. For this to be true, features for which SRL is currently dependent on such as consistent predicate-to-parse paths within automatic constituent parses must be found to exist in data such as the SSR corpus.

The predicate-argument path is defined as the number of steps up and down a parse tree (and through which NTs) which are taken to traverse the tree from the predicate (verb) to its argument. For example, the path from predicate VBP → "like" to the argument ARG0 (NP → "some kids") might be [VBP ↑ VP ↑ S ↓ NP]. As trees grow more complex, as well as more errorful (as expected for the automatic parses of verbatim speech text), the paths seen are more sparsely observed (i.e. the probability density is less concentrated at the most common paths than similar paths seen in the PropBank annotations). We thus consider two path simplifications as well:

- **compressed:** only the source, target, and root nodes are preserved in the path (so the path above becomes [VBP ↑ S ↓ NP])
- **POS class clusters:** rather than distinguish, for example, between different tenses of verbs in a path, we consider only the first letter of each NT. Thus, clustering compressed output, the new path from predicate to ARG0 becomes [V ↑ S ↓ N].

The top paths were similarly consistent regardless of whether paths are extracted from P_{r_a}, P_{v_m}, or P_{v_a} (P_{v_a} results shown in Table 4), but we see that the distributions of paths are much flatter (i.e. a greater number and total relative frequency of path types) going from manual to automatic parses and from parses of verbatim to parses of reconstructed strings.

5 Discussion

In this work, we sought to find generalizations about the underlying structure of errorful and reconstructed speech utterances, in the hopes of determining semantic-based features which can be incorporated into automatic systems identifying semantic roles in speech text as well as statistical models for reconstruction itself. We analyzed syntactic and semantic variation between original and reconstructed utterances according to manually and automatically generated parses and manually labeled semantic roles.

Argument	Path from Predicate	Freq
Predicate-Argument Paths	VBP ↑ VP ↑ S ↓ NP	4.9%
	VB ↑ VP ↑ VP ↑ S ↓ NP	3.9%
	VB ↑ VP ↓ NP	3.8%
	VBD ↑ VP ↑ S ↓ NP	2.8%
	944 more path types	84.7%
Compressed	VB ↑ S ↓ NP	7.3%
	VB ↑ VP ↓ NP	5.8%
	VBP ↑ S ↓ NP	5.3%
	VBD ↑ S ↓ NP	3.5%
	333 more path types	77.1%
POS class+ compressed	V ↑ S ↓ N	25.8%
	V ↑ V ↓ N	17.5%
	V ↑ V ↓ A	8.2%
	V ↑ V ↓ V	7.7%
	60 more path types	40.8%

Table 4: Frequent P_{v_a} predicate-argument paths

Syntactic paths from predicates to arguments were similar to those presented for WSJ data (Palmer et al., 2005), though these patterns degraded when considered for automatically parsed verbatim and errorful data. We believe that automatic models may be trained, but if entirely dependent on automatic parses of verbatim strings, an SRL-labeled resource much bigger than the SSR and perhaps even PropBank may be required.

6 Conclusions and future work

This work is an initial proof of concept that automatic semantic role labeling (SRL) of verbatim speech text may be produced in the future. This is supported by the similarity of common predicate-argument paths between this data and the PropBank WSJ annotations (Palmer et al., 2005) and the consistency of other features currently emphasized in automatic SRL work on clean text data. To automatically semantically label speech transcripts, however, is expected to require additional annotated data beyond the 3k utterances annotated for SRL included in the SSR corpus, though it may be adequate for initial adaptation studies.

This new ground work using available corpora to model speaker errors may lead to new intelligent feature design for automatic systems for shallow semantic labeling and speech reconstruction.

Acknowledgments

Support for this work was provided by NSF PIRE Grant No. OISE-0530118. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the supporting agency.

References

Eugene Charniak. 1999. A maximum-entropy-inspired parser. In *Proceedings of the Annual Meeting of the North American Association for Computational Linguistics*.

Christopher Cieri, Stephanie Strassel, Mohamed Maamouri, Shudong Huang, James Fiumara, David Graff, Kevin Walker, and Mark Liberman. 2004. Linguistic resource creation and distribution for EARS. In *Rich Transcription Fall Workshop*.

Erin Fitzgerald and Frederick Jelinek. 2008. Linguistic resources for reconstructing spontaneous speech text. In *Proceedings of the Language Resources and Evaluation Conference*.

Erin Fitzgerald, Keith Hall, and Frederick Jelinek. 2009. Reconstructing false start errors in spontaneous speech text. In *Proceedings of the Annual Meeting of the European Association for Computational Linguistics*.

Erin Fitzgerald. 2009. *Reconstructing Spontaneous Speech*. Ph.D. thesis, The Johns Hopkins University.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

Daniel Gildea and Martha Palmer. 2002. The necessity of parsing for predicate argument recognition. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.

Mary Harper, Bonnie Dorr, John Hale, Brian Roark, Izhak Shafran, Matthew Lease, Yang Liu, Matthew Snover, Lisa Yung, Anna Krasnyanskaya, and Robin Stewart. 2005. Structural metadata and parsing speech. Technical report, JHU Language Engineering Workshop.

Zellig S. Harris. 1957. Co-occurrence and transformation in linguistic structure. *Language*, 33:283–340.

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

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James Martin, and Dan Jurafsky. 2004. Shallow semantic parsing using support vector machines. In *Proceedings of the Human Language Technology Conference/North American chapter of the Association of Computational Linguistics (HLT/NAACL)*, Boston, MA.

Sameer Pradhan, James Martin, and Wayne Ward. 2008. Towards robust semantic role labeling. *Computational Linguistics*, 34(2):289–310.

Elizabeth Shriberg. 1994. *Preliminaries to a Theory of Speech Disfluencies*. Ph.D. thesis, University of California, Berkeley.

George K. Zipf. 1949. *Human Behavior and the Principle of Least-Effort*. Addison-Wesley.