Lang Resources & Evaluation (2010) 44:387–419
DOI 10.1007/s10579-010-9120-1

The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue

Sasha Calhoun · Jean Carletta · Jason M. Brenier · Neil Mayo · Dan Jurafsky · Mark Steedman · David Beaver

Published online: 11 May 2010

Abstract This paper describes a recently completed common resource for the study of spoken discourse, the NXT-format Switchboard Corpus. Switchboard is a long-standing corpus of telephone conversations (Godfrey et al. in SWITCHBOARD: Telephone speech corpus for research and development. In Proceedings of ICASSP-92, pp. 517–520, 1992). We have brought together transcriptions with existing annotations for syntax, disfluency, speech acts, animacy, information status, coreference, and prosody; along with substantial new annotations of focus/contrast, more prosody, syllables and phones. The combined corpus uses the format of the NITE XML Toolkit, which allows these annotations to be browsed and searched as a coherent set (Carletta et al. in Lang Resour Eval J 39(4):313–334, 2005). The resulting corpus is a rich resource for the investigation of the linguistic features of dialogue and how they interact. As well as describing the corpus itself, we discuss our approach to overcoming issues involved in such a data integration project, relevant to both users of the corpus and others in the language resource community undertaking similar projects.

S. Calhoun (✉)
School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland, UK
e-mail: Sasha.Calhoun@ed.ac.uk

J. Carletta · N. Mayo · M. Steedman
School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK

J. M. Brenier
Nuance Communications, Inc., Sunnyvale, CA, USA

D. Jurafsky
Department of Linguistics, Stanford University, Stanford, CA, USA

D. Beaver
Department of Linguistics, University of Texas at Austin, Austin, TX, USA

Keywords Linguistic annotation · Language resources · Discourse · Prosody · Semantics · Spoken dialogue

1 Introduction

Corpora that have been augmented with rich linguistic annotation are becoming ever more important for developing and testing theories of language. These range from detailed phonetics, such as the use of phonetic annotations to study lenition and other properties of speech production (Bell et al. 2003; Johnson 2004; Aylett and Turk 2004), to the most abstract syntax, such as the use of syntactic treebanks to study facts about information structure (Michaelis and Francis 2004) or syntactic variation (Bresnan et al. 2007). Because recording and transcribing language is expensive, corpora that are made available with transcription often attract further kinds of annotation. For instance, the Switchboard Corpus of telephone conversations (Godfrey et al. 1992), has been transcribed at the word level and annotated with parts-of-speech and syntactic structure (Marcus et al. 1993), turn and utterance boundaries and disfluency labels (Taylor et al. 2003), dialogue acts (Jurafsky et al. 1997; Shriberg et al. 1998), animacy of NPs (Zaenen et al. 2004), information status (Nissim et al. 2004), and prosodic information about prominence and boundaries (Ostendorf et al. 2001).

With such a diverse range of annotations, the Switchboard Corpus had the potential to be a very valuable resource for studying relationships and interfaces between the syntactic, semantic, pragmatic, and prosodic features of spontaneous speech. For example, many experiments have suggested that the relationships between information structure and prosodic prominence (such as whether discourse-new NPs are more likely to bear pitch accent than discourse-old NPs) are complex (e.g. Terken and Hirschberg 1994; Bard et al. 2000). A corpus that marks both information structure and prosodic prominence (as well as codes for important controls like syntactic structure and disfluency) could significantly advance our understanding of this complex relation. We could ask a wide variety of other kinds of interface questions that are important in linguistics or psycholinguistics (about, for example, the relation between speech acts and syntactic structure, the link between phonetic reduction and information status, or the relationship of disfluency and information status).

Unfortunately, the existing state of the Switchboard Corpus did not allow any of these questions to be asked. This is because these annotations were all produced in different formats by different research groups; worse, they were attached to two different underlying word transcripts of the conversations. Some of the annotations were made on the original transcript or the slightly modified *Treebank3* transcript (Marcus et al. 1993; Taylor et al. 2003), while others were made on the later, corrected *MS-State* transcript (Deshmukh et al. 1998; Harkins 2003). Moreover, they are not all available from one source. This made it very difficult to use any pair of them in combination, much less the entire set, both in terms of the time needed to do the work and the level of technical skill required. We have overcome this

difficulty by integrating all the existing Switchboard annotations into one coherent data set in the format of the NITE XML Toolkit (NXT, Carletta et al. 2005). Integrating these annotations was complicated because it required us to resolve multiple transcripts and unify different segmentations, but the resulting data has much more value than the set of separate component parts. In addition, we have added annotations for two key linguistic features of dialogue, focus/contrast and prosody; as well as syllable and phone information. These new variables, along with the wide variety of annotations already combined into the corpus, make the NXT-format Switchboard Corpus a rich resource for linguistic, psycholinguistic and computational linguistic research.

More documentation about the NXT-format Switchboard Corpus is on the corpus website (http://groups.inf.ed.ac.uk/switchboard/). The corpus has been released by the Linguistic Data Consortium (catalog number LDC2009T26, http://www.ldc.upenn.edu/) under a Creative Commons NonCommercial Share Alike license (http://www.creativecommons.org/). The Creative Commons licensing, which is similar to that for open source software, is intended to encourage users not only to use the corpus, but to offer any further annotations they make for community use. LDC has developed a separate license offering commercial terms.

We first briefly explain why the NITE XML Toolkit is the best choice for representing this data all together, and then describe each of the annotation layers and how they are represented in the NXT framework. We then show one example of a research question that can easily be investigated using the NXT-format Switchboard Corpus that would be difficult using the annotations separately. We discuss some of the more complex issues which arose in the conversion of legacy resources, particularly those issues that will be relevant for users of the final corpus who are familiar with the original format of one or more annotation. Finally, we discuss the lessons learnt about building this kind of resource generally.

2 Why the NITE XML Toolkit?

In a corpus with as many annotations as Switchboard, it is important for all of them to be in one coherent format, preferably within a framework that can be used to validate the data, read and search it, and browse it in end user tools. There are several such frameworks available, such as TIGER (Brants et al. 2002), annotation graphs (Bird and Liberman 2001), ATLAS (Laprun et al. 2002), and MMAX2 (Müller and Strube 2006). For this corpus, we chose the NITE XML Toolkit (NXT, Carletta et al. 2005).

We chose NXT for several reasons. First and foremost, of the frameworks available, only MMAX2, ATLAS, and NXT represent both temporal and explicit structural relationships among annotations. Although annotation graphs, for instance, do represent annotations as a graph structure, the semantics of edges does not cover properties like dominance (i.e. parent/child relationships). This means that such properties must be encoded within the edge labels, with no support given in software for their interpretation. NXT is more flexible in the structural relationships that it can represent than MMAX2, which uses independent stand-off layers that point to the

same base layer but cannot refer to each other. NXT allows not just more complex relationships, but also independent non-congruent structural annotations, i.e. crossing brackets. The Switchboard Corpus did not need these for the current syntactic annotation because it was originally in Penn Treebank format (Taylor et al. 2003), which does not allow for them, but they are useful for other annotations, as well as for future development. NXT also allows type-checking for long-distance dependencies, which makes checking for consistency much easier than in the original Treebank format (see Sect. 5.2). Further, NXT has more complete handling of signals, including a data handling API that makes it easier to write programs that process the data, and has the advantage of being open source. ATLAS is even more flexible in its data representation than NXT, especially with regard to pointing into signals, but its implementation is unfortunately incomplete.

In addition to its treatment of linguistic structure, NXT also has several other desirable properties. Because it separates annotations into multiple files, different people can create unrelated annotations at the same time without any additional work to merge their output afterward. Structural dominance (i.e. a parent-child relationship) is represented using XML dominance within a single file and using a particular kind of stand-off link for dominance that crosses file boundaries, making it easier to apply conventional XML processing techniques to the data. NXT also comes with a range of end user graphical interfaces for common tasks as well as libraries that can be used to write new ones efficiently. For example, there is a utility which allows users to display conversations one at a time to test queries (see Sect. 4): portions of the text returned by each query are highlighted, so that users do not have to work directly with the XML (e.g. see http://groups.inf.ed.ac.uk/switchboard/start.html). NXT also provides methods for validating that data conforms to the defined storage format. This is an important functionality that is often overlooked. Finally, NXT has an active and growing user community that has already exercised its full range of capabilities, particularly since its adoption for the popular AMI Meeting Corpus (Carletta et al. 2006).

3 The NXT-format Switchboard Corpus: annotations

The Switchboard Corpus (Godfrey et al. 1992) was collected at Texas Instruments in 1990–1991 and was released by the Linguistic Data Consortium in 1992–1993 and then again, with some errors fixed, in 1997. This 1997 "Switchboard 1 Release 2" Corpus contains recordings of about 2,400 conversations between 543 speakers of American English. Speakers chose topics of interest (e.g., cars, recycling) from a predetermined list, and were connected to each other automatically by a robotic switchboard operator. Conversations were thus between strangers. Conversations ranged in length from one and a half to ten minutes, averaging six and a half minutes. The corpus totaled roughly three million words. This original release was also transcribed, broken into turns, and diarized (labeling speakers as A and B). The corpus was then slightly improved and released as part of the Penn Treebank3 Switchboard Corpus (see details in Sect. 3.3). The NXT-format Switchboard Corpus

includes 642 of the 650 conversations from the Penn Treebank3 syntactic release. NXT Switchboard therefore includes just over 830,000 words.

Below, we begin by describing data representation within the NXT framework. We then briefly describe each layer of annotation in the NXT Switchboard, including the original annotation and how it is represented in NXT. We give more details on the *kontrast* (focus/contrast) and prosody annotations, as these have not been published elsewhere.

3.1 NXT framework

NXT models corpus data as a set of 'observations', in this case the Switchboard conversations, which are associated with one or more 'signals', here the stereo audio files. NXT allows the corpus designer to specify a 'metadata' file that describes the intended structure of a corpus; the metadata effectively combines definitions equivalent to a set of schemas for the data files with catalogue information explaining where the files can be found. The metadata file organizes annotations into multiple 'layers' that form descriptions of the corpus. For instance, typically, a transcription layer will contain tags for words, non-word vocalizations, and maybe pauses and punctuation. The designer can specify that a layer should be stored in its own file, or build up 'codings' that contain several layers, each of which hierarchically decomposes the one above it. Structural dominance is represented straightforwardly as either XML dominance, if the parent and child are in the same file, or using a 'stand-off' link notated at the parent node that indicates where to find each out-of-file child. In the data model, all children for a given node must be drawn from the same layer, and any path drawn by following only child links must not contain a cycle. This structure covers most requirements and represents a reasonable trade-off between flexibility and processing efficiency. For where it is insufficient, there is another type of stand-off link, the 'pointer', which is more flexible but incurs higher processing costs.

3.2 Transcriptions: terminals and phonwords

Underlying all the annotations we will describe are the string of words that constitute the orthographic transcript for each conversation. Unfortunately, it turns out that there were two distinct orthographic transcripts for the existing corpus, both of which had been substantially annotated. The first is the 1997 re-release of the orthographic transcript of Switchboard, the *Switchboard-1 Release 2* transcript, (Godfrey and Holliman 1997), cleaned up from the original 1993 Switchboard release. This Switchboard-1 Release 2 transcript was then used as the base for the slightly improved transcript that was included (with other annotations to be described below) in the LDC's *Treebank3* release Marcus et al. (1999). It is this version which we have used in our corpus. To avoid ambiguity, in the rest of this paper we will refer to it as the *Treebank3* transcript.

Because the Treebank3 transcript contained errors and was not time-aligned with the speech signals (Graff and Bird 2000), the Institute for Signal and Information

Processing at Mississippi State University ran a clean-up project which hand-checked and corrected the transcript of the 1126 Treebank conversations. They also produced word alignments, showing, for each transcript word, its start and end times in the audio file; word times were determined automatically, with partial manual corrections (see Deshmukh et al. 1998; Harkins 2003). We refer to the resulting time-aligned transcript as the *MS-State* transcript.

Since both the Treebank3 and MS-State transcripts had been enriched with distinct annotations, we included both transcripts separately in our corpus, using an NXT *pointer* to link equivalent words in the two versions. Section 5.1 describes the method used to create the alignment between the two transcriptions. We refer to the words from the Treebank3 transcript as *words* and the words from the MS-State transcript as *phonwords*, since the MS-State transcript words have start and end times in the audio file and hence are slightly more phonetically grounded. The double inclusion does result in redundancy, but has the advantage of retaining the internal consistency of prior annotations. For the most part, the MS-State transcription is more accurate than the Treebank3, so the other option would have been to attach all of the annotations that were derived from the Treebank transcription to the MS-State transcription and discard the original Treebank transcription. However, attaching the Treebank annotations exactly as they are would have made the resource difficult for the end-user to interpret. For instance, where the MS-State transcription adds words to the original, the syntactic annotation would appear inconsistent. On the other hand, creating new annotations to cover the changed portions of the transcription would have been time-consuming for little gain and would have greatly complicated the relationship between the NXT-format data and the original.

Figure 1 shows our solution diagrammatically. As can be seen, where there are differences in the representation of a word in the two transcripts (e.g. in the treatment of contractions like *doesn't*), one Treebank3 'word' may link to more than one MS-State 'phonword', or vice versa.

An extract of the XML representation of 'words' and 'phonwords' is given below (*doesn't* from Fig. 1). (Note that NXT has a number of graphical interfaces so that users do not have to work directly with the XML, see Sect. 4.) Each word is an

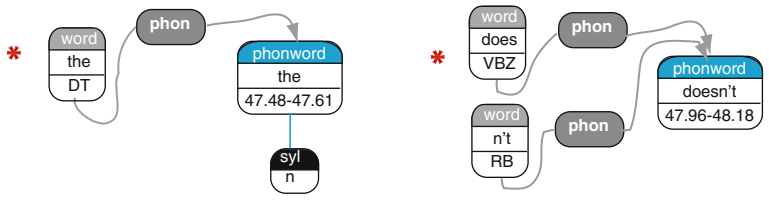


Fig. 1 Representation of the MS-State and Treebank3 Switchboard transcripts in NXT. Words in the Treebank3 transcript are represented by 'word' elements in one NXT layer, while those in the MS-State transcript are represented by 'phonword' elements in an independent layer. Representations of the same word in the two transcripts are linked by an NXT pointer labeled 'phon'. In some cases, such as contractions, words are tokenized differently in the two transcripts, so there may be multiple 'words' pointing at a 'phonword' or vice versa. Note that the star (*) shows that this structure is the expansion of the abbreviated word/phonword structure shown in Fig. 4

XML element with a unique 'nite:id', and a number of attributes, including in this case the start and end times ('nite:start' and 'nite:end'), orthography ('orth'), and part-of-speech type ('pos') for 'words'. The relationship between the elements is shown by a 'nite:pointer' on the 'word', the 'href' attribute of this pointer shows the file and 'nite:id' of the corresponding 'phonword'. All XML examples are taken from the utterance used in Fig. 4 (see Sect. 4). The file names are given bottom right, pointer relationships are demonstrated by the dashed lines, ellipses mark omitted parts of the files, and some attributes are not shown.

(1)

```
... <word nite:id="s15_56" nite:start="47.96" nite:end="n/a" pos="VBZ" orth="does">
      <nite:pointer role="phon" href="sw2708.A.phonwords.xml#id(ms16A_pw69)" />
    </word>
    <word nite:id="s15_57" nite:start="n/a" nite:end="48.18" pos="RB" orth="n't">
      <nite:pointer role="phon" href="sw2708.A.phonwords.xml#id(ms16A_pw69)" />
    </word> ...
```

sw2708A.terminals.xml

```
... <phonword nite:id="ms16A_pw69" nite:start="47.96" nite:end="48.18"
      stressProfile="pn" orth="doesn't" > ... </phonword> ...
```

sw2708A.phonwords.xml

With this approach, it is possible to use the two transcriptions independently or to traverse between them. For convenience, even though only the MS-State transcription contained timings in the original, we have copied timings over to the corresponding words from the Treebank3 transcription. NXT then automatically percolates these timings up through the discourse annotations based on the Treebank3 transcription.

3.3 Treebank: utterance boundaries, syntax, and disfluencies

We drew syntactic and disfluency annotations from the Penn Treebank Project (Marcus et al. 1993). The Penn Treebank3 release of Switchboard included annotations on 1126 of the Switchboard conversations. As we mentioned earlier, the Switchboard Release 2 transcripts had been diarized (divided into turns, each one labeled with A and B speakers). The Treebank3 release in addition segmented each turn into utterances, added part-of-speech tags on each word, and annotated each utterance for disfluencies (Meteer and Taylor 1995; Taylor et al. 2003).

The 'utterance' unit in the Treebank3 Switchboard release is a sentence-like chunk that was called a 'slash unit' in the original labeling manual (Meteer and Taylor 1995), and will be parsed as an S in the parse trees described below. The following example shows three utterances, distinguished by slashes. Notice that certain discourse markers or continuers (like *right*, and *yeah*) are segmented as utterances, and that full sentential clauses with conjunctions like *and* are often segmented off as well:

> Right, / well, when my kids were little we did have a set / and I did watch a lot of Sesame Street and a lot of Electric Company.

Each word in each utterance is part-of-speech tagged with a tag from the Penn Treebank tagset defined in Table 1.

Disfluencies (also called 'repairs') were annotated following Shriberg (1994). Figure 2 shows the structure of a repair, consisting of a *reparandum* (the 'replaced'

Table 1 NXT Word Part-of-Speech (pos) Values (from Treebank)

BES	*'s* as form of *BE*	PRP$	Possessive pronoun
CC	Coordinating conjunction	RB	Adverb
CD	Cardinal number	RBR	Adverb, comparative
DT	Determiner	RP	Particle
EX	Existential *there*	TO	Infinitival *to*
IN	Preposition/ subordinating conjunction	UH	Interjection, filler, discourse marker
JJ	Adjective	VB	Verb, base form
JJR	Adjective, comparative	VBD	Verb, past tense
JJS	Adjective, superlative	VBG	Verb, gerund/ present participle
MD	Modal	VBN	Verb, past participle
NN	Noun, singular or mass	VBP	Verb, non-3rd ps. sing. present
NNP	Proper noun, singular	VBZ	Verb, 3rd ps. sing. present
NNPS	Proper noun, plural	WDT	*Wh*-determiner
NNS	Noun, plural	WP	*Wh*-pronoun
PDT	Predeterminer	WRB	*Wh*-adverb
POS	Possessive ending	XX	Partial word, POS unclear
PRP	Personal pronoun		

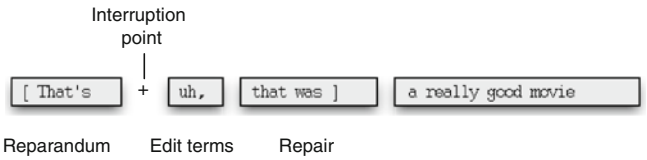


Fig. 2 The reparandum begins with a left square bracket '[' and ends with a '+'. The repair follows the (optional) edit phase after the '+' and ends with a right square bracket ']'

words), followed by an optional edit term like *uh* or *you know*, followed by the repair; see Meteer and Taylor (1995), Taylor et al. (2003).

Finally, the Treebank3 release of Switchboard also included 650 conversations (a subset of the 1126) with full syntactic parse trees. 642 of these are included in the NXT-format Switchboard Corpus release; the remaining 8 were excluded because of difficulties in processing them. The phrase level categories used in the Treebank syntactic parse trees for Switchboard are shown in Table 2. Note that the set of phrase level categories in Table 2 includes tags for the interruption point (IP), reparandum (RM), and restart/repair (RS) components of disfluencies. Long distance dependencies marked in the Treebank are represented by 'movement' elements in NXT, which show links between traces and antecedents. Syntactic phrases are also optionally marked with grammatical function tags (surface subject, logical subject) as well as semantic role tags like direction, location, manner, purpose, and time; these function tags are shown in Table 3.

In summary, the following syntactic and disfluency features are included in the NXT-format Switchboard Corpus based on the Treebank3 transcript:

Table 2 NXT Non-Terminal (nt) Category (cat) Values (from Treebank)

ADVP	Adverb Phrase	RM	Reparandum in disfluency
CONJP	Conjunction Phrase	RS	Restart after disfluency
EDITED	Reparandum in disfluency	S	Simple declarative clause
FRAG	Fragment	SBAR	Clause introduced by a (possibly empty) subordinating conjunction
INTJ	Interjection, for words tagged UH	SBARQ	Direct question introduced by a *wh*-word or *wh*-phrase
IP	Interruption point in disfluency	SQ	Inverted *yes/no* question, or main clause of a *wh*-question
NAC	Not a constituent	TYPO	Speech Error
NP	Noun Phrase	UCP	Unlike Coordinated Phrase
PP	Prepositional Phrase	VP	Verb Phrase
PRN	Parenthetical	WHADVP	*Wh*-Adverb Phrase
PRT	Particle, for words tagged RP	WHNP	*Wh*-Noun Phrase
QP	Quantifier Phrase	X	Unknown, uncertain or unbracketable

Table 3 NXT Non-Terminal (nt) Sub-Category (subcat) Values (from Treebank)

ADV	Adverbial (other than ADVP or PP)	PRP	Purpose or reason
DIR	Direction	PRP,TPC	Topicalised purpose or reason
IMP	Imperative	PUT	Locative complement of *put*
LOC	Locative	SBJ	Surface subject
LOC,PRD	Locative predicate	SBJ,UNF	Unfinished Surface Subject
MNR	Manner	SEZ	Reported speech
NOM	Nominal (on relatives and gerunds)	TMP	Temporal
NOM,TPC	Topicalised Nominal	TMP,UNF	Unfinished Temporal
PRD	Predicate (other than VP)	TPC	Topicalised
PRD,PRP	Purpose or reason predicate	UNF	Unfinished
PRD,UNF	Unfinished Predicate		

Part of speech: Penn Treebank part-of-speech (as an attribute on the *terminals*).
Turns: Syntactic sentences grouped by conversation turns and diarized (speaker A or B).
Utterances: Utterance boundaries (as the units on which *dialogue acts* are marked).
Syntax: Penn Treebank syntactic categories (Marcus et al. 1993; Taylor et al. 2003).
Movement (Long distance dependencies): Links between *traces* and *antecedents* as co-indexed in the Treebank. For example, in "What book$_i$ did you buy t$_i$?", *what book* is the antecedent of the trace, *t*.
Disfluency: Treebank disfluency coding, including *reparanda* (hesitations or false starts), *interruption points*, and *repairs*, e.g. "[the-]$_{reparandum}$ [the government]$_{repair}$".

An extract of the XML representation of 'syntax' and 'movement' is given in (2), 'turns' in (3) and 'disfluency' in (4), using the same format as (1) above (note that pointer relationships are shown by dashed lines, and child relationships by dotted lines). The antecedent in each 'movement' element is identified by a 'source' pointer, and the trace by a 'target' pointer. The syntactic category ('cat') and subcategory ('subcat') of non-terminals ('nt') are attributes. Note that turns have as children whole syntactic parses, which can include multiple clauses (in this case starting before and ending after the extract in Fig. 4). Disfluencies have two child elements, a 'reparandum' and a 'repair', each of which has a 'word' child.

(2)

```
... <movement nite:id="sw2708.reference.N4006A9" label="*">
      <nite:pointer role="source" href="sw2708.A.syntax.xml#id(s15_551)" />
      <nite:pointer role="target" href="sw2708.A.terminals.xml#id(s15_59)" />
    </movement> ...
```

sw2708A.movement.xml

```
... <nt nite:id="s15_551" nite:start="47.48" nite:end="47.96" cat="NP" subcat="SBJ">
      <nite:child href="sw2708.A.terminals.xml#id(s15_53)" />
      <nite:child href="sw2708.A.terminals.xml#id(s15_55)" /> </nt> ...
```

sw2708A.syntax.xml

```
... <word nite:id="s15_53" orth="the"> ... </word>
    <word nite:id="s15_55" orth="government"> ... </word> ...
    <trace nite:id="s15_59" /> ...
```

sw2708A.terminals.xml

(3)

```
... <turn nite:id="t9" nite:start="34.48" nite:end="48.96">
      <nite:child href="sw2708.A.syntax.xml#id(s13)..id(s15)" /> </turn> ...
```

sw2708A.turns.xml

```
... <parse nite:id="s15">
      <nt nite:id="s15_500" cat="S"> ... </nt> ... </parse> ...
```

sw2708A.syntax.xml

(4)

```
... <disfluency nite:id="sw2708.disf.5">
      <reparandum nite:id="sw2708.disf.5.reparandum">
        <nite:child href="sw2708.A.terminals.xml#id(s15_47)" /> </reparandum>
      <repair nite:id="sw2708.disf.5.repair">
        <nite:child href="sw2708.A.terminals.xml#id(s15_53)" /> </repair>
    </disfluency> ...
```

sw2708A.disfluency.xml

```
... <word nite:id="s15_47" orth="the"> ... </word> ...
    <word nite:id="s15_53" orth="the"> ... </word> ...
```

sw2708A.terminals.xml

3.4 Dialogue acts

Dialogue acts are categories of utterances much like speech acts, but drawing more on natural conversational phenomena, for example representing various acts of grounding such as backchannel responses, appreciations, and answers to questions. Jurafsky et al. (1997) annotated each utterance (slash-unit) in these same 1126 Switchboard conversations for dialogue acts using a new tagset they called SWBD-DAMSL. They used a large set of combinable tags resulting in 220 combination tags, which they then clustered into 42 dialogue act tags shown in Table 4 (as many tags were very infrequent, similar less frequent tags were clustered together, see Jurafsky et al. 1998). Both the SWBD-DAMSL tag names and the given NXT glosses are included in the data.

As we'll discuss later, the dialogue act transcripts don't exactly match the standard Penn Treebank3 transcripts, because Jurafsky et al. (1997) annotated an early version of the Penn Treebank3 transcript, after the LDC had done the utterance

Table 4 NXT Dialogue Act (da) Type Values

NXT	SWBD-DAMSL	Description	Example
abandon	%-	Adandoned or Turn-Exit	*So, -/*
acknowledge	bk	Response Acknowledgment	*Oh, okay.*
affirm	na,ny^e	Affirmative non-yes answers	*It is.*
agree	aa	Agree/Accept	*That's exactly it.*
ans_dispref	arp,nd	Dispreferred answers	*Well, not so much that.*
answer	no	Other answers	*I don't know.*
apology	fa	Apology	*I'm sorry.*
apprec	ba	Appreciation	*I can imagine.*
backchannel	b	Backchannel	*Uh-huh.*
backchannel_q	bh	Backchannel as question	*Is that right?*
close	fc	Conventional-closing	*It was nice talking to you.*
commit	oo,cc,co	Offers, Options & Commits	*I'll have to check that out.*
completion	^2	Collaborative Completion	*or not.*
decl_q	qw^d	Declarative Wh-Question	*You are what kind of buff?*
directive	ad	Action-directive	*Why don't you go first*
downplay	bd	Downplayer	*That's all right.*
excluded	@	Excluded - bad segmentation	-
hedge	h	Hedge	*Well, I don't know.*
hold	^h	Hold before response	*I'm drawing a blank.*
maybe	aap/am	Maybe/Accept-part	*Something like that.*
neg	ng,nn^e	Negative non-no answers	*Uh, not a whole lot.*
no	nn	No answers	*No.*
open	fp	Conventional-opening	*How are you?*
open_q	qo	Open-Question	*How about you?*
opinion	sv	Statement-opinion	*I think it's great.*
or	qrr	Or-Clause	*or is it more of a company?*
other	o,fo,bc by,fw	Other	*I tell you what.*
quote	^q	Quotation	*[I said] "Okay, fine"*
reject	ar	Reject	*Well, no.*
repeat	b^m	Repeat-phrase	*Oh, fajitas.*
repeat_q	br	Signal-non-understanding	*Excuse me?*
rhet_q	qh	Rhetorical-Questions	*Who has time?*
self_talk	t1	Self-Talk	*What is his name?*
statement	sd	Statement-non-opinion	*He's about five months old.*
sum	bf	Summarize/Reformulate	*So you travel a lot.*
tag_q	^g	Tag-Question	*Right?*
thank	ft	Thanking	*Hey thanks a lot.*
third_pty	t3	3rd-party-talk	*Katy, I'm on the phone.*
uninterp	%	Uninterpretable	*But, uh, yeah.*
wh_q	qw	Wh-Question	*Well, how old are you?*
yes	ny	Yes answers	*Yes.*

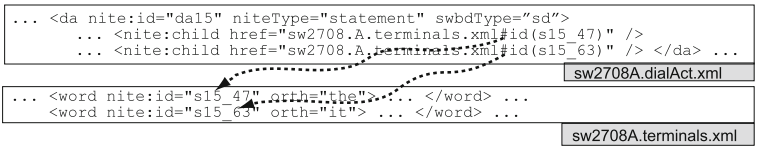
Table 4 continued

NXT	SWBD-DAMSL	Description	Example
yn_decl_q	qy^d	Declarative Yes-No-Question	*You just needed a majority?*
yn_q	qy	Yes-No-Question	*Is that what you do?*

segmentation, but in parallel with LDC's parsing of the corpus. Some corrections to mistranscriptions in both projects meant that the transcripts for the Treebank3 release and the Jurafsky et al. (1997) corpus have minor word differences.

In summary, **dialogue acts**, e.g. *statement, question*, are included in the NXT-format Switchboard Corpus based on the Treebank3 transcript. An extract of the XML representation of dialogue acts ('da') is given below. The dialogue act type in NXT is given in the attribute 'niteType', and the original SWBD-DAMSL type in the attribute 'swbdType'. Note that this dialogue act has more children than are shown, for space reasons we only give the first and last words in the utterance from Fig. 4.

(5)



3.5 Markables: animacy, information status and coreference

The 642 conversations in the Treebank3 included in the NXT-format Switchboard Corpus were further annotated for animacy (Zaenen et al. 2004) and 147 for information status (Nissim et al. 2004). As animacy and information status are properties of entities, only NPs and pronouns were marked. Disfluent speech and locative, directional, and adverbial NPs were excluded.

Animacy annotation captures the inherent accessibility of entities. Entities were marked according to widely used categories of animacy that make up an 'animacy scale', as shown in Table 5 and further described in Zaenen et al. (2004).

Information status annotation captures the accessibility of entities in a discourse, drawing on the well-known hierarchy of Prince (1992). NPs that had been previously mentioned, along with generic pronouns, were classified as *old*. NPs which had not been mentioned but were generally known or inferable were *med* (mediated). NPs which had not been mentioned and were not mediated were *new* (see Table 6). Old and mediated entities could be further classified according to a subtype, which specified how they got their old or mediated status, e.g. identity, event, situation; see Tables 7 and 8 (for more details see Nissim et al. 2004). For old entities, a *co-reference* link was also marked between references to the same entity, specifying the *anaphor* and the *antecedent*.

In summary, the following features of NPs are included in the NXT-format Switchboard Corpus based on the Treebank3 transcript:

Table 5 NXT Markable Animacy Values (from Zaenen et al. 2004)

human	Refers to one or more humans; this includes imaginary entities that are presented as human, e.g. gods, elves, ghosts
org	Collectivities of humans when displaying some degree of group identity
animal	Non-human animates, including viruses and bacteria
mac	Intelligent machines, such as computers or robots
veh	Vehicles
place	Nominals that "refer to a place as a place", e.g. *at my house*
time	Expressions referring to periods of time
concrete	"Prototypical" concrete objects or substances, e.g. body parts; excluded are things like air, voice, wind and other intangibles
nonconc	The default category; used for events, and anything else that is not prototypically concrete but clearly inanimate
oanim	Coder unsure of animacy status
mix_an	Mixed animacy status
anim_uncoded	Animacy status uncoded

Table 6 NXT Markable Info Status Values (from Nissim et al. 2004)

old	Entity has been mentioned before, or is generic (see examples in Table 7)
med	Not mentioned before, but can be inferred from previous discourse or general knowledge (see examples in Table 8)
new	Newly mentioned and not inferable
status-uncoded	Information status uncoded

Table 7 NXT Markable *Old* Info StatusType Values (from Nissim et al. 2004)

ident	Anaphoric reference to a previously mentioned entity, e.g. *I met* M. **He**'*s a nice guy*
relative	Relative pronoun
generic	Generic pronoun, e.g. *in holland* **they** *put mayo on chips*
ident_generic	Generic possessive pronoun, e.g. *in holland they put mayo on* **their** *chips*
general	*I* and *you*
event	Reference to a previously mentioned VP, e.g. *I like* going to the mountains. *Yeah, I like* **it** *too*
none	Sub-category not specified

Animacy: Coding of NPs for animacy status, e.g. *human, animal, non-concrete* (as an attribute on the *markables*).
Information Status: Coding of NPs as *old, mediated* or *new*, plus sub-types of *old* and *mediated* (as an attribute on the *markables*).
Coreference: Links between each *anaphor* (i.e. NP marked as *old-identity* and its *antecedent* (i.e. its previous mention in a conversation).

An extract of the XML representation of 'markables' and 'coreference' is shown in (6). The 'markable' element has attributes showing the 'animacy' type, information 'status' and information status sub-type ('statustype'). Markables point at NPs (an 'nt', note the child of this 'nt' is the word *it*, as shown in (5). Coreference

Table 8 NXT Markable *Mediated* Info StatusType Values (from Nissim et al. 2004)

bound	Bound pronoun, e.g. everyone *likes* **his** *job*
general	Generally known, e.g. *the sun*
event	Relates to a previously mentioned VP, e.g. *We were* traveling around Yucatan, *and* **the bus** *was really full*
aggregation	Reference to previously mentioned co-ordinated NPs, e.g. John... Ann... **they**
func_value	Refers to the value of a previously mentioned function, e.g. **in**... centigrade ... *between* **zero** and **ten** *it's cold*
set	Subset, superset, or member of the same set as a previously mentioned entity
part	Part-whole relation for physical objects, both intra- and inter-phrasal, e.g. *when I come* home *...my dog greets me at* **the door**
poss	Intra-phrasal possessive relation (pre- and post-nominal) that is not *part*
situation	Part of a situation set up by a previous entity, e.g. capital punishment... **the exact specifications**
none	Sub-category not specified

elements have two pointers, to the 'anaphor' and the 'antecedent', both of which are 'markables' (note only one is shown here).

(6)

```
... <link nite:id="sw2708.coreference.6">
      <nite:pointer role="anaphor" href="sw2708.A.markable.xml#id(sw2708.markable.43)"/>
      <nite:pointer role="antecedent" href="sw2708.A.markable.xml#id(sw2708.markable.37)"/>
    </link> ...
```

sw2708A.coreference.xml

```
... <markable nite:id="sw2708.markable.43" animacy="nonconc" status="old"
     statustype="ident">
       <nite:pointer role="at" href="sw2708.A.syntax.xml#id(s15_560)" /> </markable> ...
```

sw2708A.markable.xml

```
... <nt nite:id="s15_560">
      <nite:child href="sw2708.A.terminals.xml#id(s15_63)" /> </nt> ...
```

sw2708A.syntax.xml

3.6 Kontrast and triggers

A total of 145 conversations from the set annotated for information status have also been annotated for kontrast (focus/contrast). While focus-marking has been extensively discussed in both the semantics literature (e.g. Halliday 1968; Rooth 1992; Steedman 2000) and the intonational phonology literature (e.g. Pierrehumbert and Hirschberg 1990; Selkirk 1995; Ladd 2008), there have been few attempts to annotate focus in corpora. Most existing studies use a rather restrictive definition of focus, as being either new information or an explicit contrast in the context (e.g. Nakatani et al. 1995; Hedberg and Sosa 2001; Zhang et al. 2006) (though see Buráňová et al. 2000). We have used a much broader notion of focus, based on the widely-accepted Alternative Semantics definition (Rooth 1992). We call focus under this definition *kontrast*, following Vallduví and Vilkuna (1998), to distinguish our usage from other definitions of focus in the literature and the common usage of *contrast* which might imply only explicit contrasts. To our knowledge, there have been no other attempts to annotate foci using this definition in unrestricted speech; so our scheme is novel.

Table 9 NXT Kontrast Type Values

correction	Corrects or clarifies another word or NP just used by either speaker, e.g. *now are you sure they're* **hyacinths**, *because that is a bulb*.
contrastive	Intended to contrast with another word mentioned in the context, e.g. *I have got some in the* **backyard** *that bloomed blue... I would have liked those in the* **front**. A *trigger* marks the link between *backyard* and *front*.
subset	Highlights one member of a more general set that has been mentioned and is a current topic, e.g. *this woman owns* **three day cares**... *she had to open* **the second one** *up because her waiting list was a year long!* Again, a *trigger* links the set (*day cares*) and the subset (*the second one*).
adverbial	A focus-sensitive adverb, i.e. *only*, *even*, *always*, *especially*, *just*, *also* or *too* is used to highlight the word, and not another in a plausible set, e.g. *(A) I thought [Michael J Fox] was crummy in 'The Hard Way'. (B) I didn't* **even** *like the* **previews** *on that*. A *trigger* linked the adverb and kontrast.
answer	Fills an open proposition set up in the context such that it would make sense if only that word or phrase were spoken, e.g. *(A) [these] blooms... I'm not sure what they are... they come in all different colors ... (B) I'm going to bet you that is a* **LILY**.
other	Clearly kontrastive, but not one of the other types, used sparingly.

Table 10 Distribution of kontrast types at the word and NP level (frequencies exclude non-applicables)

Type	Word	NP	Freq	Type	Word	NP	Freq (%)
Contrastive	6823	1885	7.8%	Answer	196	116	0.3
Other	6166	1544	6.9%	Correction	169	54	0.2
Subset	5037	2273	6.6%	Background	91856	n/a	82.7
Adverbial	1798	160	1.8%	Non-Applicable	13325	n/a	–
Total	124440	6962	111115				

Annotators identified words or NPs which were "salient with an implication that this salience is in comparison or contrast to other related words or NPs explicitly or implicitly evoked in the context"; that is, explicitly using the alternative semantics definition. However, annotators did not mark kontrast directly. Instead, words or NPs were marked according to their *kontrast types* (based on Rooth 1992), see Table 9; with all other words being *background*. A decision tree was used where more than one category applied; this ranked kontrast types according to their perceived salience in different sentential contexts (for full details see Calhoun 2005, 2006, Chap. 5). It was felt that this indirect approach was more natural and immediately comprehensible to both annotators and other eventual users of the corpus. In certain categories, annotators also marked a *trigger* link between the kontrast and the word that motivated its category assignment (see Table 9). Table 10 shows the overall distribution of kontrast annotation types.

Only certain words were annotated, i.e. nouns, verbs, adjectives, adverbs, pronouns and demonstratives in full sentences. This was done to improve the efficiency of the annotation, as it was thought these words would be most likely to be genuinely kontrastive. Further, annotators were instructed to mark false starts,

hesitations and idiomatic phrases such as "in fact" or "you know" as *non-applicable*; as Alternative Semantics theory did not seem to cover such cases. Annotators could listen to the conversation (but not see acoustic information). We felt this was important to identify actual kontrasts intended by the speaker, rather than all potential kontrasts, given the highly ambiguous nature of spoken spontaneous discourse.

Annotations were done by two post-graduate linguistics students at the University of Edinburgh. Annotators were given fairly extensive training, after which they reported that they understood and felt confident about their task. Agreement was measured on three conversations at different stages of the project using the kappa statistic (Siegel and Castellan 1988). In all cases, annotators did two passes of each conversation, i.e. they checked their own annotations once. "Blind" kappa agreement, i.e. without discussion, over all kontrast types was $\kappa = 0.67$, and $\kappa = 0.67$ for the binary distinction between kontrast and background ($k = 2$, $n = 3{,}494$). Given the level of confidence of the annotators, this was lower than hoped, but is not unusual. Being a new task, it is difficult to know what a "good" score is. Therefore we also measured an "agreed" figure: in each case where the annotators disagreed, each explained to the other the reason for the type they chose; where they could both then agree that one type was better (with reference to the guidelines and decision tree), the other changed their annotation. Both annotators were considered equals. "Agreed" kappa over all kontrast types was $\kappa = 0.85$, and $\kappa = 0.83$ for the kontrast/background distinction ($k = 2$, $n = 3494$). This could be considered an "upper bound" on annotator agreement for this task, at least using the current definitions.

Two particular sources of annotator disagreement were identified. One was caused by the varying *scope* of kontrast. Annotators were able to mark kontrast at the word or NP level. It was decided it would be too difficult to maintain consistency if the size of kontrast elements were unrestricted. Therefore, when the kontrast appeared to be larger than the NP, annotators were instructed to mark the word or words which sounded most salient. This led to conflicts about salience which did not actually stem from disagreement about kontrast status. This issue is difficult to resolve, and in fact Carletta (1996) notes that segmentation is one of the most challenging aspects of discourse annotation, and may make certain tasks inherently more uncertain than others, such as clause boundary identification. We also found disagreement where one or more kontrast type plausibly applied, but one analysis or the other was not noticed by one of the annotators for the "blind" comparison, or then accepted for the "agreed" comparison. For research purposes so far considered, we consider such discrepancies in annotation minor provided that each such case was annotated as some sort of kontrast (i.e. not background). More problematic were the fairly common disagreements between *other* and *background*. Overall, we decided it was better to keep the category, because of the many cases which were clearly kontrastive, but did not fit in one of the other types. The annotators' difficulty does vindicate our decision not to annotate kontrast per se, however. In general, the annotations were reasonably successful, given the lack of precedent for annotating focus in spontaneous English conversation. Further development of such a standard will want to look again at the issue of kontrast scope and the status of *other*.

In summary, the following features of content words are included in the NXT-format Switchboard Corpus based on the Treebank3 transcript:

Kontrast: Coding of words for whether they have a salient alternative in the context (*kontrast*), or not (*background*). Kontrast was marked according to a number of types, e.g. *contrastive, subset, answer*.
Trigger: Links certain kontrasts to the word(s) that motivated their marking.

An extract of the XML representation of 'kontrasts' and 'triggers' are shown in (8). The 'type' of the kontrast, and the 'level' at which it was marked (word or NP), are attributes of the 'kontrast' element. 'Triggers' had two pointers, a 'referent', which pointed at the main kontrast, and a 'trigger', which pointed at the element which motivated the kontrast marking (not shown here). For ease of comprehension, the context of this example is given in (7). The conversation is about who should pay for the prison system, the child of `sw2708.kontrast.48` is *government* (cf. (2)), and the child of `sw2708.kontrast.42` is *business* (XML links to words not shown below).

(7) *they're talking about having it [the prison system] as a* business... *so... the* government *doesn't have to deal with it.*

(8)

```
... <trigger nite:id="sw2708.trigger.3">
      <nite:pointer href="sw2708.kontrast.xml#id(sw2708.kontrast.48)" role="referent" />
      <nite:pointer href="sw2708.kontrast.xml#id(sw2708.kontrast.42)" role="trigger" />
    </trigger> ...
```

sw2708A.trigger.xml

```
... <kontrast level="word" nite:id="sw2708.kontrast.48" type="contrastive">
      <nite:child href="sw2708.A.terminals.xml#id(s15_55)" /> </kontrast> ...
```

sw2708A.kontrast.xml

3.7 Prosody: accents, phrases and breaks

The NXT-format Switchboard Corpus includes simplified ToBI prosody annotation of 45 complete conversations (definitions below). It also includes conversations which were annotated by us according to ToBI-like standards with some modifications: 18 of these are complete conversations, while in 13 further only sentential clauses containing kontrastive words were annotated (the majority). There are some other existing prosodic annotations for Switchboard Corpus data (Taylor 2000; Yoon et al. 2004), however, we have not tried to include these as they are of isolated utterances, which are not useful for dialogue analysis. Below we describe each prosodic annotation set, and then how these are represented in NXT Switchboard.

The 45 conversations with simplified ToBI markup were annotated for prosody by Ostendorf et al. (2001), based on the MS-State transcript. These are identified as the "UW [University of Washington] prosody" annotation set. Accents and phrase breaks were annotated using simplified To(nes) and B(reaks) I(ndices) (ToBI) labels (Beckman and Hirschberg 1999). Annotators labeled a break index tier, identifying 0, 1, 1p, 2, 2p, 3, 4 and X breaks (broadly higher indices show greater break strength, 'p' is disfluent, 'X' is uncertain break index); and a tone tier, labeling L-, H- and X (low, high and uncertain) phrase accents at 3 breaks, as well as L%, H% and X% boundary tones at 4 breaks. At 3 breaks, they could also use !H- phrase

accents for a mid-range pitch fall after a high accent ('!' indicates downstepped). A question mark after the tone label indicated the annotator was unsure of their classification. In an accent tier, accents were identified using a *, or *? for a weak accent. Tonal pitch accent type was not labeled.

Our prosody annotation scheme was also based on the ToBI standards. We have, however, made certain changes to concentrate on features which were most important for our research questions, and were useful generally (for full details see Brenier and Calhoun 2006; Calhoun 2006, Chap. 5). As well as marking the presence of accents, in our scheme, one accent in each fluent phrase was identified as *nuclear*. As far as we are aware, this feature is unique among available corpora. However, nuclear accents have long been claimed to have important properties distinct from other accents (Crystal 1969; Liberman 1975; Ladd 2008). The nuclear accent is a compulsory part of a well-formed phrase, while the presence of other accents varies depending on rhythm, length and emphasis. Further, it has often been claimed that nuclear accents, not accents in general, mark focus (see Ladd 2008). The nuclear accent was defined as the most *structurally* prominent, normally the right-most one, not necessarily the most *phonetically* prominent (Ladd 2008). After some discussion and practice, the annotator was able to use this concept effectively. There were a few difficult cases, particularly in phrases with an early emphatic high accent and a later, lower nuclear accent. We therefore introduced a *pre-nuclear* marker for the first accent in such cases; this was used rarely, however. Non-nuclear accents could be marked as either *weak* or *full*, to cover cases in which the cues to accenting were mixed. Tonal accent type was not marked.

Rather than marking a break index after every word, as in ToBI, in our scheme, words were grouped into prosodic phrases, and then each phrase was marked as being one of four types. Unified phrases were felt to be more useful for the investigation of the relationship between discourse functions and phrasing than sequential break indices. Fluent phrases could be either *minor* (ending in ToBI break index 3) or *major* (ending in break index 4). As in ToBI, the distinction was based on the perceived degree of disjuncture, as well as the tonal movement at the boundary. Phrase breaks that sounded disfluent, e.g. cut-offs before restarts, repetitions or hesitations, were marked *disfluent* (equivalent to ToBI 1p and 2p). Short phrases containing only discourse fillers, e.g. *um*, *you know*, with no tonal movement, were marked *backchannel* (ToBI X). An example *Praat* textgrid and acoustic display (Boersma and Weenink 2006) showing the prosody annotation are given in Fig. 3. As can be seen, 'accents' mark points of prosodic prominence (marked primarily with intensity by this speaker), 'phrases' prosodic groupings of words, and 'breaks' the degree of juncture after each word.

In all, 31 Switchboard conversations have been annotated for prosody using our scheme. Of these, we annotated 13 from scratch (designated as the "Ed [University of Edinburgh] original prosody" annotation set), and only included words in sentential clauses which had also been annotated for kontrast (see above), as these were intended to form a complementary set. The remaining 18 conversations (designated as the "Ed converted" annotation set) were annotated by manually converting the annotations on conversations that had already been marked up using the annotation scheme of Ostendorf et al. (2001). This approach made use of the

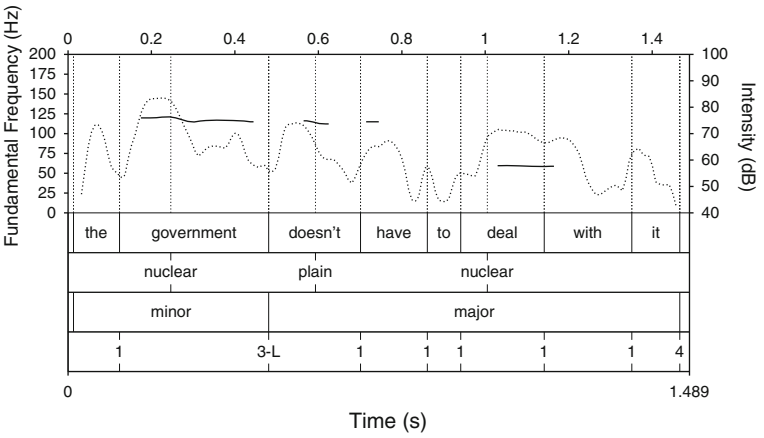


Fig. 3 Example *Praat* textgrid and acoustic display of the prosody annotation (part of the utterance from Fig. 4). The fundamental frequency track (*solid line*) and intensity curve (*dotted line*) are shown, along with the phonword transcript, the accent annotation (*accent type marked*), the phrase annotation (*type marked*), and break annotation (break index and phrase tone marked)

existing data and was more efficient than starting from scratch. As well as converting the annotations, the annotator corrected anything with which they disagreed. Unlike the 13 conversations annotated from scratch, in these conversations all words were marked, as they had been in the originals. Most of the annotations and conversions were done by a post-graduate linguistics student at the University of Edinburgh with experience using ToBI, and a small number (3) by the first author. Annotations were carried out for each conversation side separately on the MS-State transcript using *Praat* (Boersma and Weenink 2006), and then later moved to NXT format. Table 11 shows the overall distribution of accent and phrase types.

One conversation side was used to check agreement between our annotator and the first author. This comparison was "blind", i.e. the annotators had no access to each other's annotations before agreement was measured. Kappa agreement on the presence or absence of a phrase break following each word was $\kappa = 0.91$, and on phrase break type was $\kappa = 0.89$ ($k = 2,\ n = 752$). Agreement on the presence or absence of an accent, and on accent type, was $\kappa = 0.80$ ($k = 2,\ n = 752$). These

Table 11 Distribution of accent and phrase types using our annotation scheme (Ed original/Ed converted sets)

Accent	Freq (%)	*N*	Phrase	Freq (%)	*N*
Nuclear	25.8	(12322)	Minor	11.1	(5269)
Pre-nuclear	0.3	(156)	Major	15.0	(7119)
Full non-nuclear	11.1	(5340)	Disfluent	1.6	(783)
Weak non-nuclear	3.6	(1710)	Backchannel	1.8	(871)
Unaccented	59.1	(28207)	No break	70.5	(33537)
Total		47735	Total		47579

scores are high enough that the research community would accept them without question. They are also commensurate with those reported for previous ToBI annotation projects (Pitrelli et al. 1994; Yoon et al. 2004), suggesting the changes made in our scheme were successful. There is little difference in kappa for all types versus presence/absence (±), showing good discrimination between types.

All prosody annotations just described are represented in NXT using three elements: accents, phrases and breaks. Because of the differences in the way our three sets of source files (UW, Ed original and Ed converted) were annotated, there are slight differences in how these elements are generated for each set. However, we have generated a full set of all three elements for all conversations annotated. In this way, the NXT representation retains most of the information in the originals, while the entire set of 76 conversations annotated for prosody can be searched as a set. Because of the considerable annotator time and expense needed, it was not possible to annotate all conversations according to both prosody annotation schemes, so that all the source material was uniform; should that even be desirable.

For all three sets of annotations, accents are represented at the time marked in the original annotation in the NXT representation. All accents have a strength attribute, *weak* versus *full* (*? versus * in the UW annotation); the Ed original and converted conversations also have a type attribute, *nuclear* versus *plain* (see Table 12). An NXT pointer marks the word that accent associates with: in the Ed sets, this was marked by the annotators; for the UW set, the word association was derived automatically from the word timings. The two annotation schemes marked prosodic boundaries differently: in the Ed scheme as phrases, i.e. words grouped into prosodic phrases; in the UW scheme as breaks, i.e. the degree of juncture after each word. As each contained slightly different information, and different users may find either breaks or phrases more useful, it was decided to include both breaks and phrases in the NXT representation (see Table 12). It is anticipated that users will use one or the other. For the Ed sets, phrases were derived directly from the manual annotation. For the UW set, phrases were generated automatically using the locations of higher-level break indices (3, 4, 2p, 3p or X). As the information about break indices in the UW annotations was richer than the Ed break index information, breaks were generated from the original UW annotations where these existed. Breaks point at the word they fall after, and include the break index, and associated phrase tone and boundary tone, if there are any. For the Ed original

Table 12 NXT Accent, Phrase and Break Attributes and Values

Element	Attribute	Values
accent	strength	weak, full
	type (Ed only)	nuclear, plain
phrase	type	minor, major, disfluent, backchannel
break	index (Ed)	3, 4, 2p, X
	index (UW)	Full ToBI break index (1-4, p, X...)
	phraseTone (UW only)	L, H, !H, X (+ ? variant)
	boundaryTone (UW only)	L, H, X (+ ? variant)

conversations only, breaks were derived automatically from phrases, so only 2p, 3, 4, and X breaks are marked; and there are no phrase or boundary tones.

In summary, the following prosodic features are included in the NXT-format Switchboard Corpus based on the MS-State transcript:

Accent: Pitch accents (*weak* or *full*), marked at the time of the peak and associated with words. Word association was marked manually for the Ed sets, automatically for the UW set. Accent type is given for the Ed sets (*nuclear* or *plain*).

Phrase: Groupings of words into prosodic phrases by type (*minor*, *major*, *disfluent*, *backchannel*); marked manually for the Ed sets, determined automatically from the manual ToBI break marking for the UW set.

Breaks: ToBI break indices, phrase and boundary tones, derived from UW annotations. For the Ed original set, generated automatically from the phrases, so only 2p, 3, 4 and X breaks marked and phrase/boundary tones not included.

An extract of the XML representation of 'accents' is shown in (9), 'phrases' in (10) and 'breaks' in (11). Accents are marked at a single point of time (usually the pitch peak), this is represented by having the 'nite:start' and 'nite:end' times the same. The accent 'strength' and 'type' are attributes. Note that accents point at 'phonwords' (MS-State transcript). Phrases have 'phonword' children, while 'breaks' point at the 'phonword' they follow (note the break in (11) points at the 'phonword' *government*, shown in (10)). Breaks have attributes showing the break 'index', and optionally, the 'phraseTone' and 'boundaryTone' (note (11) does not have the latter). Breaks are also marked at a single point in time (the word end), so 'nite:start' equals 'nite:end' in (11).

(9)

```
... <accent nite:id="sw2708.A.acc37.aw103" nite:start="47.70" nite:end="47.70"
      strength="full" type="nuclear">
       <nite:pointer role="at" href="sw2708.A.phonwords.xml#id(ms16A_pw68)" />
    </accent> ...
```

sw2708A.accents.xml

```
... <phonword nite:id="ms16A_pw68" orth="government"> ... </phonword> ...
```

sw2708A.phonwords.xml

(10)

```
... <phrase nite:id="sw2708A.phrase.28" nite:start="47.48" nite:end="47.96" type="minor">
      <nite:child href="sw2708.A.phonwords.xml#id(ms16A_pw67)..id(ms16A_pw68)" />
   </phrase> ...
```

sw2708A.phrases.xml

```
... <phonword nite:id="ms16A_pw67" orth="the"> ... </phonword>
    <phonword nite:id="ms16A_pw68" orth="government"> ... </phonword> ...
```

sw2708A.phonwords.xml

(11)

```
... <break nite:id="sw2708A.break.69" nite:start="47.96" nite:end="47.96"
     index="3" phraseTone="L">
       <nite:pointer href="sw2708.A.phonwords.xml#id(ms16A_pw68)" role="at" />
    </break> ...
```

sw2708A.breaks.xml

3.8 Phones and syllables

Finally, automatically derived phone and syllable annotation layers have been added for all 642 conversations in the NXT-format Switchboard Corpus, based on the

MS-State transcript. Although a small set of Switchboard utterances have been hand-transcribed phonetically (Greenberg et al. 1996), this set was drawn from independent utterances from many different conversations, and hence did not comprise any complete conversations. Thus we have automatically derived phone and syllable identity and timings for our entire corpus. Firstly, using the whole Switchboard Corpus, acoustic triphone models were trained with *Sonic*, the University of Colorado continuous speech recognition system (Pellom 2001). Next, these models were used to generate forced phone alignments from the MS-State transcript's word start and end times. The resulting phone sequences were automatically parsed into syllables using NIST syllabification software (Fisher 1997), and each syllable was assigned a stress level (*primary*, *secondary*, or *unstressed*) using the CMU pronunciation dictionary (Weide 1998). Automatic phone and syllable alignment technology is fairly mature, so this information could be derived with reasonable efficacy.

In summary, the following sub-lexical features are included in the NXT-format Switchboard Corpus based on the MS-State transcript:

Syllables: Automatically derived syllables, including stress information (*primary*, *secondary* or *none*).
Phones: Automatically derived phones, includes the start and end time for each phone.

An extract of the XML representation of 'syllables' and 'phones' ('ph' elements) is shown below. Timing information is included on the 'phone' and 'phonword' levels, this can be used to get the timing on the 'syllable' level. The 'stress' on each syllable is an attribute (n = no stress, p = primary, s = secondary). This is used to generate the 'stressProfile' attribute on the 'phonword', i.e. a list of the stress information for all its syllable children [also see (1)].

(12)

```
... <phonword nite:id="ms16A_pw67" stressProfile="n" orth="the">
      <nite:child href="sw2708.A.syllables.xml#id(ms67A_sy1)" /> </phonword> ...
                                                              sw2708A.phonwords.xml
... <syllable nite:id="ms67A_sy1" stress="n">
      <nite:child href="sw2708.A.phones.xml#id(ms102A_ph1)..id(ms102A_ph2)" />
    </syllable> ...
                                                              sw2708A.syllables.xml
... <ph nite:id="ms102A_ph1" nite:start="47.48" nite:end="47.53">dh</ph>
    <ph nite:id="ms102A_ph2" nite:start="47.53" nite:end="47.61">ah</ph> ...
                                                              sw2708A.phones.xml
```

In addition to these layers, there are NXT 'corpus resource' files representing information about the topics and speakers for the dialogues.

4 Corpus overview and use

All of the annotations described above are represented in different layers of annotation in NXT, and the relationships between them faithfully represented using NXT parent/child and pointer relationships (see Sect. 3.1). Here, we give an

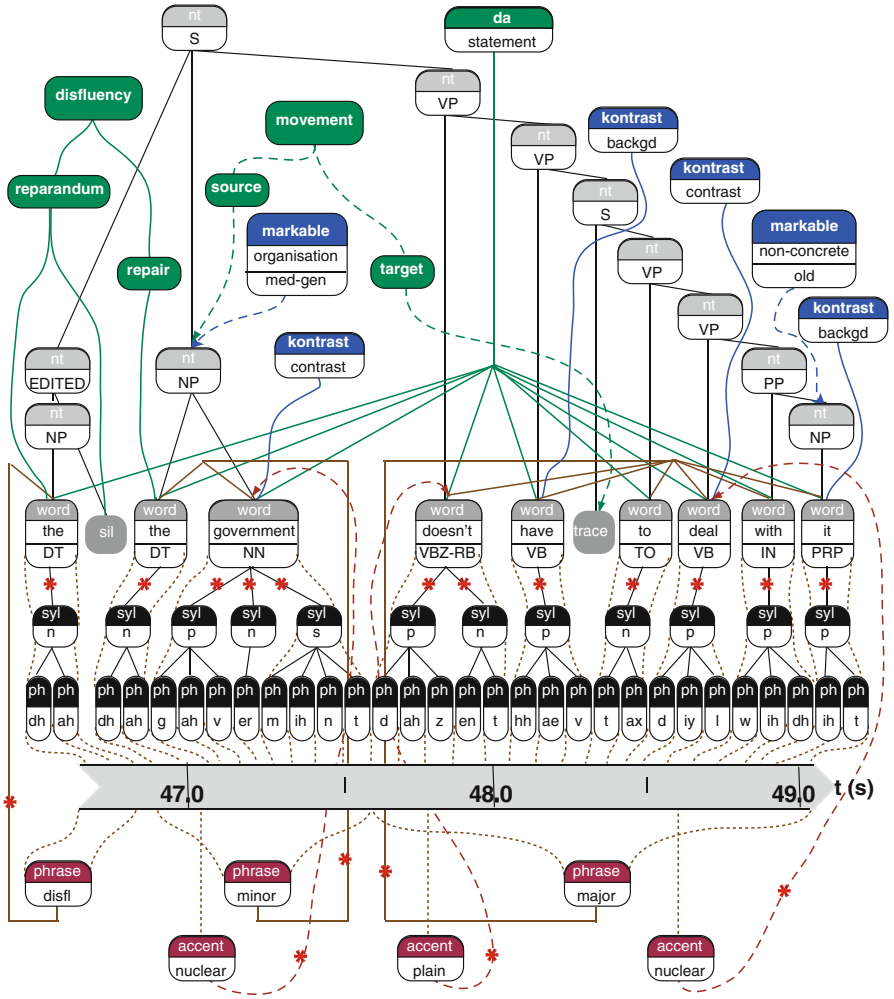


Fig. 4 Overview of annotations from a small sample of the Switchboard Corpus as represented in the NXT data model. Individual nodes may have multiple attributes; for simplicity, we show just the values of the most important ones. Parent/child relationships are shown by *solid lines*, pointers by *dashed lines*. Note that where the relationship between a word and another element is marked with a star (*), this word in fact points at a phonword, which is directly linked to the other element (see Fig. 1). Turn, coreference, trigger and break codings are not shown. See http://groups.inf.ed.ac.uk/switchboard/ for further details

overview of the overall structure of the corpus, and a small example of how it can be queried using NXT tools.

Figure 4 shows a simplified example of the resulting structure drawn from the corpus (note this uses the same example as in the preceding sections). Not all annotations cover the entire corpus. Figure 5 shows the number of Switchboard conversations with each type of annotation.

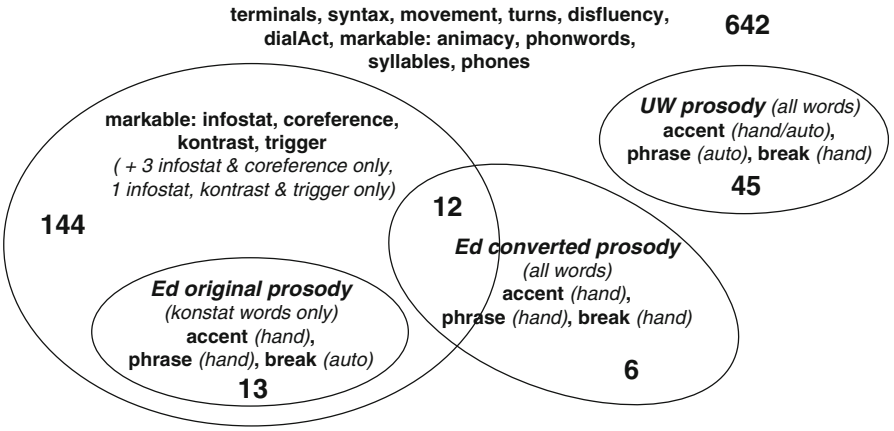


Fig. 5 Diagram showing the number of conversations in the NXT-format Switchboard Corpus release with the different layers of annotation (see http://groups.inf.ed.ac.uk/switchboard/coverage.html for a list of which conversations are in each set). Note that the information status and kontrast annotations were intended to cover the same subset. However, there are four anomalous files that are missing one or more of the relevant layers. The prosody files (accent, phrase and breaks) were generated in slightly different ways, either directly from manual annotations (hand) or automatically from other annotations (auto), and cover different numbers of words, either all words or only those also annotated for kontrast status (konstat only). These annotations, therefore, are listed according to their source: Edinburgh (Ed) original, Ed converted or University of Washington (UW). Further details of the prosody annotations are discussed in Sect. 3.7

One principal advantage of our corpus is that it allows all of the annotations to be searched as a set. NXT's query language (NQL, Carletta et al. 2005) allows search for n-tuples of nodes from the graph structure where the variable bindings can be constrained by type and filtered using boolean combinations of conditions that express either restrictions on a node's attribute values or temporal and structural relationships between pairs of nodes.

For instance, it is often said that in English 'new' information is accented while 'given' information is not. Evidence from controlled experiments and restricted domain speech (such as map tasks) shows that the situation is more complex than this (e.g. Terken and Hirschberg 1994; Bard et al. 2000). However, to our knowledge, this has not been tested in a large scale corpus of unrestricted speech. This can be done easily in the NXT-format Switchboard Corpus. The analyst must first identify a query that specifies how to pull out pairs of 'markables' coded for information status and accents that go together (note that NXT provides graphical interfaces to assist with this, see http://groups.inf.ed.ac.uk/switchboard/start.html). These variables are not directly related in the NXT corpus structure (see Fig. 4), so the query must specify the path between these two layers of annotation, i.e. markables point at noun phrases that contain some orthographically transcribed word(s), and the accent points at the corresponding phonetic word ('phonword'):

```
($m markable)($nt nt)($w word)($pw phonword)($a accent):
                        \\RETURN MARKABLE, NT, WORD, PHONWORD and ACCENT 5-tuples
    ($m > $nt) &&
                        \\WHERE THE MARKABLE POINTS AT A NON-TERMINAL (NT)
        ($nt ^ $w) &&
                        \\AND THAT NT IS THE PARENT OF A WORD
            ($w > $pw) &&
                        \\AND THAT WORD POINTS AT A PHONWORD
                ($a > $pw)
                        \\AND AN ACCENT ALSO POINTS AT THAT PHONWORD
```

This kind of query allows users to retrieve properties of words contained in annotations attached to both of the transcripts (e.g. markables and prosody). However, the results returned will fail to cover unaligned segments of one or other transcription. The NXT-format Switchboard Corpus thus allows all the same investigations as the components from which it was created, but makes it easier to do research that uses several kinds of annotation, as well as to add new ones.

The analyst would then use one of NXT's command line utilities to extract the data from the corpus (see the NXT website for details, http://groups.inf.ed.ac.uk/nxt/nxtdoc/doccommandlinetools.xml). The utility chosen depends on the form of output the analyst requires, e.g. XML elements or plain text. For example, the 'FunctionQuery' utility pulls out specified attributes of entities that were matched in the query, such as the orthography of the word (`$w@orth`), the markable's information status (`$m@status`), and the accent type (`$a@type`), and returns them in plain text tab-delimited format, like the following (for details on the code to run this utility see http://groups.inf.ed.ac.uk/switchboard/start.html):

```
      $w@orth       $m@status     $a@type
...   business      med           nuclear
      government    med           nuclear
      I             old           plain     ...
```

A similar analysis would then need to be performed to obtain information about unaccented cases (note an extra condition, that the 'phonword' has a 'phrase' parent, would need to be added to ensure the word is prosodically annotated, see Sect. 3.7). A collation of the results of applying this analysis to the 25 conversations annotated for both information status and accenting (including accent type, i.e. nuclear/non-nuclear) is shown in Table 13. It seems that the majority of both 'new' and 'old' words in NPs are unaccented. 'Mediated' and 'New' words are more likely to carry nuclear accents, but not non-nuclear accents. A full research project looking at this issue may also wish to look at sub-types of 'old' and 'mediated', how long since the last mention for 'old' words, or filter the results by the syntactic position of the word; all of which are possible with the NXT-format Switchboard Corpus.

Table 13 Accent Type by Information Status: example query results from NXT-format Switchboard Corpus

	Accent Type						
Info Status	None		Non-Nuclear		Nuclear		Total
Old	3528	(66.5%)	883	(16.7%)	891	(16.8%)	5302
Med	6007	(57.4%)	1530	(14.6%)	2928	(28.0%)	10465
New	2208	(54.0%)	738	(18.0%)	1145	(28.0%)	4091

5 The conversion process

The development of the NXT-format Switchboard Corpus involved the conversion of the existing corpus annotations in a variety of different formats, as described in Sect. 3, into one coherent data set. The conversion of a set of legacy resources in diverse formats into one coherent data set is always a messy process, especially when performed piecemeal over a number of years. However, we believe the process of doing this shows why using XML within the NITE framework is worthwhile, because it provides a validatable data format, GUIs and a search language. This is useful not only for end users, but during the process of data creation and annotation resolution itself. During the conversion process we identified (and were often able to fix) inconsistencies in the source data, as well as test and refine assumptions about relationships within and between annotations which are not always obvious even to its creators. The resulting XML layers could also be straight-forwardly checked for consistency and usefulness using NXT's query language.

5.1 Transcription alignment

The process of aligning the Treebank3 and MS-State transcript, i.e. creating a pointer relationship between equivalent words in the two transcripts, involved a number of stages and necessarily remains imperfect. The first stage of the alignment process involved matching words where the two transcripts were the same. After this process, 6.9% of the Treebank3 words and 7.7% of the MS-State words were unaligned. The difficulty at this stage was determining which of the words were unaligned due to actual differences between the transcripts, and which of these words should be treated as equivalent. We assumed a match if the words were the same, apart from differences in capitalization or some other minor point of orthography. We also assumed a match for different representations of filled pauses (e.g. *uh* vs. *ah-huh*), disfluent or unfinished words (e.g. *gov-* vs. *govern-[ment]*), and non-verbal cues like laughter. We created some equivalences which mapped more than one word to a phonword, and vice versa. These involve contractions (e.g. MS-State *don't* vs. Treebank3 *don't*), the form of reduced words (e.g. *wanna* vs. *want to*), and acronyms (e.g. MS-State *IBM* vs. Treebank3 *I B M*). Finally, we aligned corresponding gaps in the two transcripts if they only involved one or two words,

even if the words were different, as manual checks suggested the timing information was correct. The transcription alignment process described here leaves 0.5% of the Treebank3 words and 2.2% of the MS-State words unaligned. To the best of our knowledge, these unaligned words represent genuine differences between the two transcripts, such as where one transcript has a word between two aligned words that the other does not.

Finally, the NXT version of the Treebank3 transcription sometimes differs in the speaker to whom a transcription segment is attributed. There were some swapped speakers for entire Switchboard conversations, which were fixed by revised speaker tables which we took into account (Graff and Bird 2000). In addition to these errors, however, by comparing to the MS-State transcription, we found that there were some additional swapped sentences within individual conversations as well. We used the MS-State transcription as the definitive source for information about speaker identity, and therefore corrected the speaker attribution for these swapped sentences in the Treebank3 transcript.

The lesson here is that transcription changes, even fairly minor edits, cause major difficulties for corpus representation once that corpus has multiple dependent annotations. Transcription changes are inevitable for a living corpus, but it is very expensive and time-consuming, if it is possible at all, to update all annotations to reflect the new underlying transcript. Our parallel representation is faithful, but it is a pretty uneconomical way to store a corpus. For corpora designed entirely within the NXT framework, it is possible to specify a version for each file of annotations and the dependencies among them, providing a partial solution to this problem—if an annotation relies on an old version of the transcription or of some other annotation, then NXT can be instructed to use that old version. This allows the corpus users to migrate annotations to newer versions if that becomes scientifically necessarily, but still works in the meantime.

5.2 Conversion of Penn Treebank release

Carletta et al. (2004) reports on the process by which the Penn Treebank release of transcription, syntax, and disfluency annotation was converted into a precursor of the current NXT format. The main difference between the two NXT formats was that the original did not separate transcriptions and annotations for the two speakers into different files corresponding to NXT ‘agents’; although agents were part of the NXT data model at the time, separation would have provided no benefits to the originators of the conversion. This is because no word timings were available, and, therefore, the material could only be treated as written transcription without any link to audio. In addition, the precursor NXT format contained a flat list of disfluencies, whereas the current NXT format nests disfluencies hierarchically where appropriate. The previous version of the NXT format was used for a range of investigations that focused primarily on discourse and syntax, but this format has been superceded by the current one.

When the original version of the NXT-format Switchboard Corpus was created from the Penn Treebank release, data validation and checking backtranslations against the original revealed that some of the disfluencies were entirely missing

from the translation. This was because the originals were missing part of the markup expected; the 'EDITED' constituent, or the 'DFL' disfluency marker, or part of the bracketing. This is understandable. The disfluency markup forms a tree crossing the Penn Treebank syntax that is difficult to validate when it is stored in the same file, but easy in NXT. As part of the conversion process, but not mentioned in Carletta et al. (2004), the missing markup was inserted to make these disfluencies complete. As a result, using NXT Switchboard may result in more disfluencies than were in the original format, depending on the parsing assumptions made.

The TreeBank E_S and N_S tags, used to mark end-of-utterance and end-of-incomplete-utterance respectively, were not maintained in NXT Switchboard, since the information is recoverable from the parse trees in the NXT representation.

5.3 Conversion of dialogue acts

In NXT Switchboard, the dialogue acts draw words derived from the Treebank3 transcription as children. However, these words are not exactly the same as in the original dialogue act distribution. For the most part, it is possible to map between these two transcriptions automatically, converting, for instance, 'can not' to 'cannot'. However, the dialogue act distribution contains representations of non-words such as laughter, noise, and comments that were not in the Treebank3 transcription, but omitted Treebank3's 'MUMBLEx' and commas, used for mumbled or unintelligible speech. In addition, the dialogue act distribution sometimes contains 'slash units' in which a speaker's turn is split between two transcribed turns, with a turn from the other speaker transcribed in between. In these cases, the Treebank3 transcription and the dialogue act distribution differ in turn order; the Treebank places the incomplete turns one after the other, with the alternate speaker's turn following them. We migrated the dialogue acts to the Treebank3 transcription by allowing an act to match the Treebank's words despite these minor differences, and ordered them according to the Treebank convention.

In addition, the dialogue act distribution contains something akin to disfluency annotation that was present in the pre-Treebank transcription, but discarded in the Treebank release: annotation for asides, coordinating conjunctions, discourse markers, explicit edits, and fillers (Meteer and Taylor 1995). Although some of these might be considered superceded by the Treebank syntax, not all would be, and the results of using the original and the Treebank would not be entirely the same. We ignored this markup in our conversion. It would be a relatively simple matter to retrieve it and place it in a new, separate NXT hierarchy.

6 Discussion

As can be seen, the NXT-format Switchboard Corpus substantially develops and improves upon existing corpus resources. The NXT framework itself enables both effective representation of all existing annotations, and efficient integration of new layers of annotation. Further, the annotation set now available is unique in its coverage

of important linguistic features of dialogue, including syntax, disfluency, speech acts, animacy, information status, coreference, kontrast and prosody. NXT Switchboard is potentially of great benefit for a variety of researchers across linguistics, psychology and speech technology interested in data-driven analysis of the effect of multiple linguistic factors on spontaneous speech. To date the NXT-format corpus has been used to predict accents (Sridhar et al. 2008; Nenkova et al. 2007; Brenier et al. 2006; Calhoun 2006, ch. 6, 2010), kontrast (Badino and Clark 2008; Sridhar et al. 2008; Nenkova and Jurafsky 2007; Calhoun 2006, ch. 6, 2007, 2009) and information status (Sridhar et al. 2008; Nissim 2006). This corpus has also proved useful to investigate syntactic variation, cued by animacy and information status (Bresnan et al. 2007), complexity (Jaeger and Wasow 2005) and syntactic priming (Reitter 2008; Reitter et al. 2006; Dubey et al. 2005). The corpus is particularly well suited for this since it is fully parsed, allowing easy extraction of the relevant variant cases; and the many layers of annotation allow for much more control of potential interacting factors than is usually possible with naturally occurring speech.

Performance is always a worry when working with complex data sets. It is difficult to give a general idea of NXT's speed in running queries, because so much depends on the machine and the processing being done with the data. Clearly, for some queries, NXT is slower than other query languages, but this is because it is searching relationships that cannot be expressed in them; most languages are designed for tree-structured data models, which are easier to process, and do not include any operators for quantification. NXT holds data as a graph structure in memory, which can be a limiting factor. However, earlier issues with how this design choice scales have been addressed in recent releases. Current NXT selectively loads only the data files that are required for what the user is doing. There are very few data uses that would require all of the annotations for a conversation to be loaded at the same time. Similarly, it is very rarely necessary to load multiple conversations at once instead of merely iterating over them. NXT has been used successfully for a wide range of purposes on both this and other corpora. Our experience is that really complicated linguistic analyses need to be run in batch at the command line, but that for most queries, the response times when browsing the data are sufficient.

The history of the Switchboard Corpus shows that even a single layer of annotation for a significant amount of text or speech are useful and will be sought after by those outside the research group that created it. The generous agreement of developers of all these annotations of the Switchboard corpus to make their annotations freely available should, we hope, act as a positive example for our field. Having the Switchboard Corpus and all of the associated annotations in one consistent, searchable format clearly adds value over and above that found in the individual annotations. We have shown that it is possible to convert existing annotations of a corpus into a coherent format to get maximum use out of them. However, this is far from the ideal way to put together language resources. The resulting corpus is never going to be as good as a resource that is put together in an integrated framework in the first place, because there are losses along the way, e.g. invalid data, incompatible transcriptions, ambiguities in the documentation and missing documentation.

Ideally, multiple annotations should be planned from the beginning of a project (as for example with the AMI Meeting Corpus, Carletta et al. 2006). Unfortunately, this takes long-term, coherent planning and funding beyond the resources or aims of many research groups. A good place to start in creating mutually beneficial corpus resources is to agree on a consistent and flexible data format that can be validated and is underpinned by software, like the one that underlies NXT. We recognize that NXT in its current form lacks some of the end user tools that are required, and that it has limitations—chief of which is that it is difficult for less computationally-oriented users. On the other hand, it is hard to imagine any simpler framework that will allow the kinds of novel investigations to be done that are the point of this kind of corpus in the first place. In addition, there are many common corpus creation and annotation tasks for which using NXT is now already the easiest solution. For research communities that genuinely wish to foster data re-use and the more complex analyses this enables, using and further developing NXT will in the end be simpler, and more affordable, than doing the kind of post-hoc conversion process described here. In any case, the research community has much to gain from consolidating how it stores and processes corpus data.

As should be clear, there is plenty of scope for future work: on the tools we have developed, on the new NXT-format Switchboard Corpus, and on richly annotated integrated corpora in general, whether legacy resources, new resources, or hybrids. And while we have concentrated in this closing section on lessons learned from our resource building enterprise, it is important to stress that the immediate contribution of our research is not merely a set of observations on preferred methodology, but a set of ready-to-use research resources. We have shown why the NITE XML Toolkit is a good choice for representing complex combinations of corpus annotation data, and how the new resources described in this paper can facilitate research on issues like information structure and prosody. With these resources, researchers can perform corpus studies of interactions between disparate aspects of language operating at every level from acoustic signal to discourse structure, interactions that were previously inaccessible.

Acknowledgements This work was supported by Scottish Enterprise through the Edinburgh-Stanford Link, and via EU IST Cognitive Systems IP FP6-2004-IST-4-27657 "Paco-Plus" to Mark Steedman. Thanks to Bob Ladd, Florian Jaeger, Jonathan Kilgour, Colin Matheson and Shipra Dingare for useful discussions, advice and technical help in the development of the corpus and annotation standards; and to Joanna Keating, Joseph Arko and Hannele Nicholson for their hard work in annotating. Thanks also to the creators of existing Switchboard annotations who kindly agreed to include them in the corpus, including Joseph Piccone, Malvina Nissim, Annie Zaenen, Joan Bresnan, Mari Ostendorf and their respective colleagues. Finally, thank you to the Linguistics Data Consortium for agreeing to release the corpus under a ShareAlike licence through their website, and for their work in finalising the corpus data and permissions for release.

References

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1):31–56.

Badino, L., & Clark, R. A. (2008). Automatic labeling of contrastive word pairs from spontaneous spoken English. In *IEEE/ACL Workshop on Spoken Language Technology*, Goa, India.

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*(1), 1–22.

Beckman, M., & Hirschberg, J. (1999). The ToBI annotation conventions. http://www.ling.ohio-state.edu/∼tobi/ame_tobi/annotation_conventions.html. Accessed 9 June 2006

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*(2), 1001–1024.

Bird, S., & Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Communication, 33*(1–2), 23–60.

Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer. http://www.praat.org. Accessed 9 June 2006.

Brants, S., Dipper, S., Hansen, S., Lezius, W., & Smith, G. (2002). The TIGER Treebank. In *Proceedings of the workshop on Treebanks and linguistic theories*, Sozopol.

Brenier, J., & Calhoun, S. (2006). Switchboard prosody annotation scheme. Internal Publication, Stanford University and University of Edinburgh: http://groups.inf.ed.ac.uk/switchboard/prosody_annotation.pdf. Accessed 15 January 2008.

Brenier, J., Nenkova, A., Kothari, A., Whitton, L., Beaver, D., & Jurafsky, D. (2006). The (non)utility of linguistic features for predicting prominence in spontaneous speech. In *Proceedings of IEEE/ACL 2006 workshop on spoken language technology*, Aruba.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Buráňová, E., Hajičová, E., & Sgall, P. (2000). Tagging of very large corpora: Topic-focus articulation. In *Proceedings of COLING conference* (pp. 278–284), Saarbrücken, Germany.

Calhoun, S. (2005). Annotation scheme for discourse relations in Paraphrase Corpus. Internal Publication, University of Edinburgh: http://groups.inf.ed.ac.uk/switchboard/kontrast_guidelines.pdf. Accessed 15 January 2008.

Calhoun, S. (2006). *Information structure and the prosodic structure of English: A probabilistic relationship*. PhD thesis, University of Edinburgh.

Calhoun, S. (2007). Predicting focus through prominence structure. In *Proceedings of interspeech*. Antwerp, Belgium.

Calhoun, S. (2009). What makes a word contrastive: Prosodic, semantic and pragmatic perspectives. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Where prosody meets pragmatics: Research at the interface*, Vol. 8 of *Studies in pragmatics* (pp. 53–78). Emerald, Bingley.

Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*. Special Issue on Prosody (to appear).

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics, 22*(2), 249–254.

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraiij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, M., Post, W., Reidsma, D., & Wellner, P. (2006). The AMI Meeting Corpus: A pre-announcement. In S. Renals & S. Bengio (Eds.), *Machine learning for multimodal interaction: Second international workshop*, Vol. 3869 of *Lecture notes in computer science*. Springer.

Carletta, J., Dingare, S., Nissim, M., & Nikitina, T. (2004). Using the NITE XML toolkit on the Switchboard Corpus to study syntactic choice: A case study. In *Proceedings of LREC2004*, Lisbon, Portugal.

Carletta, J., Evert, S., Heid, U., & Kilgour, J. (2005). The NITE XML Toolkit: Data model and query language. *Language Resources and Evaluation Journal, 39*(4), 313–334.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge, UK: Cambridge University Press.

Deshmukh, N., Ganapathiraju, A., Gleeson, A., Hamaker, J., & Picone, J. (1998). Resegmentation of Switchboard. In *Proceedings of ICSLP* (pp. 1543–1546), Sydney, Australia.

Dubey, A., Sturt, P., & Keller, F. (2005). Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *HLT/EMNLP*, Vancouver, Canada.

Fisher, W. M. (1997). tsylb: NIST Syllabification Software. http://www.nist.gov/speech/tool. Accessed 9 October 2005.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92* (pp. 517–520).

Godfrey, J. J., & Holliman, E. (1997). *Switchboard-1 Release 2*. Linguistic Data Consortium, Philadelphia. Catalog #LDC97S62.

Graff, D., & Bird, S. (2000). Many uses, many annotations for large speech corpora: Switchboard and TDT as case studies. In *LREC*, Athens, Greece.

Greenberg, S., Ellis, D., & Hollenback, J. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard Corpus. In *The fourth international conference on spoken language processing* (pp. S24–S27), Philadelphia, PA.

Halliday, M. (1968). Notes on transitivity and theme in English: Part 3. *Journal of Linguistics, 4*, 179–215.

Harkins, D. (2003). Switchboard resegmentation project. http://www.cavs.msstate.edu/hse/ies/projects/switchboard. Accessed 1 February 2005.

Hedberg, N., & Sosa, J. M. (2001). The prosodic structure of topic and focus in spontaneous English dialogue. In *Topic & focus: A workshop on intonation and meaning*. University of California, Santa Barbara, July 2001. LSA Summer Institute.

Jaeger, T. F., & Wasow, T. (2005). Processing as a source of accessibility effects on variation. In *Proceedings of the 31st Berkeley Linguistics Society*.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29–54), Tokyo, Japan, 2004. The National International Institute for Japanese Language.

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., & Ess-Dykema, C. V. (1998). Switchboard discourse language modeling project report. Center for Speech and Language Processing, Johns Hopkins University, Baltimore, MD, 1998. Research Note No. 30.

Jurafsky, D., Shriberg, E., & Biasca, D. (1997). Switchboard SWBD-DAMSL Labeling Project Coder's Manual, Draft 13. Technical Report 97-02, University of Colorado Institute of Cognitive Science .

Ladd, D. R. (2008) Intonational phonology (2nd edn.). Cambridge, UK: Cambridge University Press

Laprun, C., Fiscus, J. G., Garofolo, J., & Pajot, S. (2002). A practical introduction to ATLAS. In *Proceedings of LREC*, Las Palmas, Spain.

Liberman, M. (1975). *The intonational system of English*. PhD thesis, MIT Linguistics, Cambridge, MA.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*, 313–330.

Marcus, M. P., Santorini, B., Marcinkiewicz, M. A., & Taylor, A. (1999). *Treebank-3*. Linguistic Data Consortium (LDC). Catalog #LDC99T42.

Meteer, M., & Taylor, A. (1995). Disfluency annotation stylebook for the Switchboard Corpus. Ms., Department of Computer and Information Science, University of Pennsylvania, http://www.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps. Accessed 30 September 2003.

Michaelis, L. A., & Francis, H. S. (2004). Lexical subjects and the conflation strategy. In N. Hedberg & R. Zacharski (Eds.), *Topics in the grammar-pragmatics interface: Papers in honor of Jeanette K. Gundel* (pp. 19–48), Benjamins.

Müller, C., & Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods*, English Corpus Linguistics (Vol. 3, pp. 197–214), Peter Lang.

Nakatani, C., Hirschberg, J., & Grosz, B. (1995). Discourse structure in spoken language: Studies on speech corpora. In *Working notes of the AAAI spring symposium on empirical methods in discourse interpretation and generation* (pp. 106–112), Stanford, CA.

Nenkova, A., Brenier, J., Kothari, A., Calhoun, S., Whitton, L., Beaver, D., & Jurafsky, D. (2007). To memorize or to predict: Prominence labeling in conversational speech. In *NAACL human language technology conference*, Rochester, NY.

Nenkova, A., & Jurafsky, D. (2007). Automatic detection of contrastive elements in spontaneous speech. In *IEEE workshop on automatic speech recognition and understanding (ASRU)*, Kyoto, Japan.

Nissim, M. (2006). Learning information status of discourse entities. In *Proceedings of the empirical methods in natural language processing conference*, Sydney, Australia.

Nissim, M., Dingare, S., Carletta, J., & Steedman, M. (2004). An annotation scheme for information status in dialogue. In *Fourth language resources and evaluation conference*, Lisbon, Portugal.

Ostendorf, M., Shafran, I., Shattuck-Hufnagel, S., Carmichael, L., & Byrne, W. (2001). A prosodically labeled database of spontaneous speech. In *Proceedings of the ISCA workshop on prosody in speech recognition and understanding* (pp. 119–121), Red Bank, NJ.

Pellom, B. (2001). SONIC: The University of Colorado continuous speech recognizer. Technical Report TR-CSLR-2001-01, University of Colorado at Boulder.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press, Cambridge, MA.

Pitrelli, J., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proceedings of the third international conference on spoken language processing* (Vol. 2, pp. 123–126).

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295–325). Philadelphia/Amsterdam: John Benjamins.

Reitter, D. (2008). *Context effects in language production: Models of syntactic priming in dialogue corpora*. PhD thesis, University of Edinburgh.

Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the conference of the cognitive science society* (pp. 685–690), Vancouver, Canada.

Rooth, M. (1992). A theory of focus intepretation. *Natural Language Semantics, 1*, 75–116.

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, MA & Oxford: Blackwell.

Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies*. PhD thesis, University of California at Berkeley.

Shriberg, E., Taylor, P., Bates, R., Stolcke, A., Ries, K., Jurafsky, D., Coccaro, N., Martin, R., Meteer, M., & Ess-Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech, 41*(3–4), 439–487.

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd edition). McGraw-Hill.

Sridhar, V. K. R., Nenkova, A., Narayanan, S., & Jurafsky, D. (2008). Detecting prominence in conversational speech: Pitch accent, givenness and focus. In *Speech prosody*, Campinas, Brazil.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry, 31*(4), 649–689.

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America, 107*, 1697–1714.

Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn Treebank: An overview.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: Effects of persistence of grammatical role and surface position. *Language and Speech, 37*, 125–145.

Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. *Syntax and Semantics, 29*, 79–108.

Weide, R. (1998). The Carnegie Mellon Pronouncing Dictionary [cmudict. 0.6]. Carnegie Mellon University: http://www.speech.cs.cmu.edu/cgi-bin/cmudict. Accessed 9 October 2005.

Yoon, T.-J., Chavarría, S., Cole, J., & Hasegawa-Johnson, M. (2004). Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In *Proceedings of ICSLP*, Jeju, Korea.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M., & Wasow, T. (2004). Animacy encoding in English: Why and how. In B. Webber & D. Byron (Eds.), *ACL 2004 workshop on discourse annotation* (pp. 118–125).

Zhang, T., Hasegawa-Johnson, M., & Levinson, S. (2006). Extraction of pragmatic and semantic salience from spontaneous spoken English. *Speech Communication, 48*, 437–462.