

What makes a word contrastive?

Prosodic, semantic and pragmatic perspectives *

Sasha Calhoun

University of Edinburgh, UK

September 7, 2009

Abstract

Contrast is an important feature of discourse that lies at the intersection of the prosodic, semantic and pragmatic systems. By looking what makes a word contrastive from each of these perspectives, we can shed light on central concerns in each area and how they interact. Following Rooth (1992), we take all semantic foci (or contrasts) to be theoretically contrastive, in that they invoke a presupposition of alternatives to the focused element. We claim that contrast acts as a strong constraint on the probabilistic alignment of words with metrical prosodic structure; contrasts align with nuclear accents. We further argue that speakers can exploit expectations about the interaction of these two systems in order to derive pragmatic effects. The more prominent a word than expected (given its prosodic, syntactic and discourse properties), the more likely it will be ‘contrastive’, i.e. the salience of its alternative set increases. Speakers can manipulate the nature and salience of the resultant presuppositions in order to create and draw attention to intended conversational implicatures of an utterance. We set out a range of evidence for these claims using a spoken language corpus, Switchboard. We describe substantial new annotations of contrast and prosody carried out on the corpus. We then present contrast prediction models developed using the corpus which show that contrast can be predicted by syntactic and prosodic features and their interaction, as claimed. Finally, we analyse selected extracts from the corpus, to demonstrate how contrastive interpretations are derived and their pragmatic effects.

*The research on which this work is based was completed as part of a Scottish Enterprise Edinburgh-Stanford Link grant. The author is now supported by a British Academy postdoctoral fellowship. Thanks to Bob Ladd and Mark Steedman for many useful discussions in developing the ideas presented here; also to Jean Carletta, Jason Brenier, Neil Mayo, Dan Jurafsky, David Beaver, Jonathan Kilgour, Shipra Dingare and Leonardo Badino for their efforts in constructing the corpus and extracting features. Thanks to the audience in the Prosody and Pragmatics in Spoken Language Corpora Panel at the 10th International Pragmatics Conference in Göteborg for helpful feedback on the original presentation of this paper; and the editors of this volume for useful comments on an earlier draft.

1 Introduction

Contrastiveness is a common feature of discourse which lies at the intersection of the prosodic, semantic and pragmatic systems. By examining what makes a word contrastive, we can shed light on central concerns in each area and how they interact.

In this paper, we take *contrastiveness* to mean when a speaker *intends to make salient* the contrast between a particular element in an utterance and some other entity or concept (cf. Umbach 2004). We can begin to see this in the following example, from the Switchboard corpus (described in section 3). Figure 1 shows its acoustic representation (sound file *movie*), along with accenting (nuclear (N), full non-nuclear (A) or weak non-nuclear (A?)) and prosodic phrasing (round brackets), and contrast annotation (see section 3.1).¹

(1) A1: I think the only [time] I've ever seen a sequel was 2001. 2010 was far better.

A2: **2001 was a good movie,**

A3: if you had read the book.

We will discuss the interpretation of A2 (in bold), which was said by a male speaker. The context sets up a contrast between *2001* and *2010*. This is marked by a strong pre-nuclear accent on *two* (and a weaker accent on *one*). However, there is also a contrast on *movie*. If we simply read A2, we might expect *movie* to be given, entailed by *2001* (e.g. see Allerton 1978). We would then expect the second contrast to be on *good*, implicitly contrasting with *bad* or *not as good as 2010*. In fact, the speaker contrasts *movie* and *book* (i.e. there is a nuclear accent on *movie*, but no accent on *good*). Our claim is that the speaker does this in order to introduce a contrast with different media for *2001*, e.g. *book* (which had not been relevant before). This has at least two pragmatic effects, it shifts the topic from *sequels*, and evokes his intended implicature: that *2001* is not as good as *2010* because you have to have read the book to understand *2001*.

The speaker achieves these pragmatic effects by manipulating which elements are in focus, and how salient those foci are. As we will argue, all such foci are contrastive, in that they presuppose alternatives

¹All examples discussed in this paper are from Switchboard. In this and all following examples, A or B indicates the speaker, where there is more than one utterance by the same speaker these are numbered A1, A2, etc. for ease of reference. The utterance to be analysed is in bold. The prosodic segmentation of the context is shown by the punctuation. Words in square brackets were not said by the speaker, but are included to make the meaning clear. The relevant sound files, along with Praat textgrids (Boersma & Weenink 2008), are at http://www.psy.ed.ac.uk/people/scalhoun/contrast_pragmatics/.

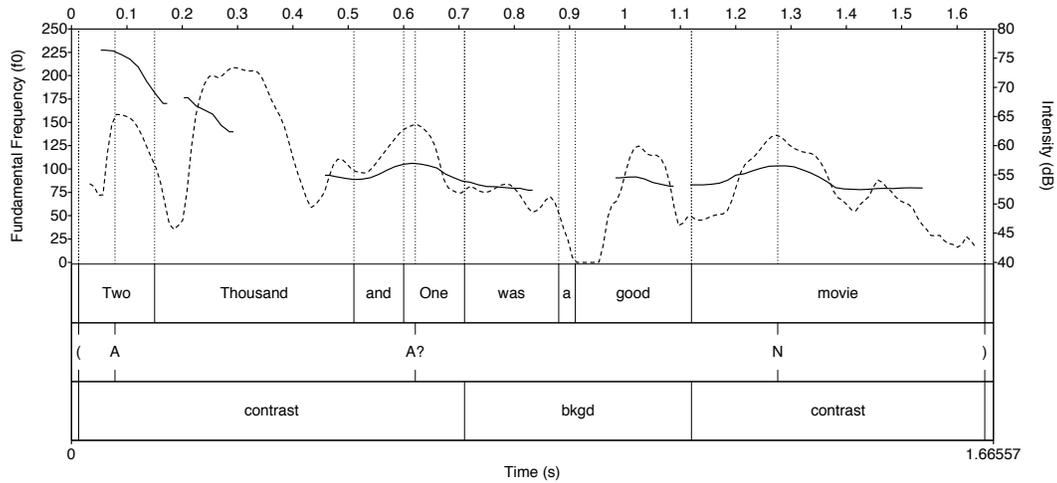


Figure 1: f_0 trace (solid line) and intensity curve (dashed line) for (1), along with the word transcript, accent type and phrasing, and contrast type annotation (see section 3.1).

(Rooth 1992). Speakers use these presuppositions to create conversational implicature. Prosody is a principal marker of focus. This manipulation therefore takes place at the interface between the prosodic system and the information structure (focus) system.

In the first part of the paper, we develop our account of how contrast is marked in English using prosodic cues within an expectation-based probabilistic framework, and discuss the implications for semantic and pragmatic interpretation. We begin by reviewing the role of contrast in semantics and pragmatics, and how contrast is marked prosodically, before outlining our approach in section 2. In the second part of the paper we set out a range of evidence for our theory by looking at contrast in a spoken language corpus, Switchboard. To date there have been few corpus studies linking prosodic prominence to discourse features like contrast. However, we think this is important, both to verify theories about this relationship developed primarily using constructed examples, and to see how it is used for pragmatic effect. In section 3, we report firstly on our annotation of both contrast and prosody in our corpus. We then present results of a study which automatically predicted the contrast status of words in the corpus. Finally, we examine extracts from the corpus to see how speakers signal contrast and the effect on pragmatic interpretation.

1.1 Contrast in Semantics and Pragmatics

In semantics, contrast is part of *information structure*, i.e. how each entity, predication, etc. in an utterance refers back to, alters and/or updates the existing discourse model (e.g. Steedman 2000). However, it is important for pragmatics as well, as speakers can exploit contrast for pragmatic effect, through the presuppositions that come from alternative sets.

The main linguistic mechanism standardly held to control information structure is *focus* (e.g. Jackendoff 1972, Rooth 1992, Selkirk 1995). Following Vallduví & Vilkuna (1998) and Steedman (2000), we take focus-related phenomena to in fact involve two separate dimensions: the ‘quantificational’ distinction between *kontrast* (K) and *background*, and the ‘organisational’ distinction between *rheme* (ρ) and *theme* (θ) (see also Kruijff-Korbayová & Steedman 2003). As Vallduví & Vilkuna (1998) argue, these two dimensions are orthogonal, capturing different aspects of information structure. We can see this in an analysis of (1):

(2) [[2001]_K] _{θ} [was a good [movie]_K] _{ρ}

We define *kontrast* according to the Rooth’s (1992) widely-used alternative semantics definition of focus (e.g. Lambrecht 1994, Truckenbrodt 1995, Vallduví & Vilkuna 1998, Steedman 2000, Büring 2006).² That is, K-marked elements introduce, in addition to the ordinary semantic meaning of a proposition, a presupposition of alternatives to the *kontrast* (his focus). *Kontrasts* are defined in relation to the theme or rheme unit that they fall in. The *scope* of the *kontrast*, i.e. which elements within the theme or rheme unit are part of the alternative set, can extend from the K-marked element itself to the whole theme or rheme unit, depending on the intended meaning (see further next section). Elements in each theme or rheme unit that are not *kontrastive* are *backgrounded*. For example, the *kontrast* interpretation of the rheme (ρ) in (2) introduces a presupposition of alternatives to the type of media of *2001*, while *is good* is backgrounded:

(3) $\text{is}(2001, \text{good movie}) \supset \neg[\text{is}(2001, \text{good book}) \mid \text{is}(2001, \text{good stage play}) \mid \dots]$

²Note that while many works adopt Rooth’s definition of focus, some maintain that it only applies to ‘contrastive’ or ‘identificational’ focus, as opposed to ‘informational’ focus (e.g. Lambrecht 1994). In our approach, these are not opposites, since rhematicity (akin to informational focus) is orthogonal to *kontrast*. Rooth himself does not use the theme/rheme dimension. However, he does analyse utterances such as (1) as involving two *kontrasts* (his foci), without discussing the functional distinction between them. We believe, therefore, that our approach is not inconsistent with his.

The contrasts on *2001* and *movie* have different functions. *2001* is thematic (θ), or topical, linking this utterance to the preceding one; while *movie* is rhematic (ρ), marking the ‘assertion’ or (contextually) ‘new’ information. All rhemes contain a contrast, though themes can contain contrasts too, as is shown here (Steedman 2000).³ The limit of the theme or rheme unit is determined by ‘projection’ from the position of the contrast to the phrase structure around it (also called focus projection), as we will see in the next section. While we believe distinguishing these two dimensions of information structure greatly clarifies and enhances the analysis of focus-related phenomena, our main purpose is to explore the semantic and pragmatic effects of K(ontrast)-marking; as this is evidently the most important for an analysis of contrast. Therefore, further arguments for the two-dimensional analysis of focus laid out here, and the role of the theme/rheme dimension, must be left to other work (see Kruijff-Korbayová & Steedman 2003, Gundel & Fretheim 2004, Calhoun 2006, Calhoun 2008).

Under our analysis, even “out-of-the-blue” (all rheme) utterances contain a contrast (cf. Bolinger 1965, p. 106). For example, if (1-A2) had instead been said at the opening of a conversation, then the contrast would probably have scope over the whole utterance (so called ‘broad focus’, Ladd 1980), as follows (note that the *scope* of the contrast refers to the elements in the theme/rheme unit which have alternatives, and *projection* to the limit of the theme/rheme unit):

- (4) [[2001 was a good movie]_K]_ρ ≡
 is(2001,good movie) ⊃ ¬[is(Eyes Wide Shut,bad movie) | coming(mother,today) | ...]

Such an account neatly circumvents arguments about the distinction between ‘normal’ and ‘contrastive’ focus by unifying their formal description (e.g. see Gussenhoven 1983). All contrasts are *theoretically contrastive*. By this we mean that the fact of contrast-marking introduces a presupposition of alternatives to the contrast, i.e. a contrast. However, there are substantial differences in how that presupposition is resolved. In our argument below, we make an important distinction between *kontrast*, defined above, and *contrastiveness*, when the speaker *intends* the alternatives to the contrast to be salient. This can be related to the scope of the contrast: the broader the scope, the less well-defined the alternative set, and therefore the less likely a clear contrastive interpretation. The key factor, though, is really the *availability* of the alternative set. The more salient and clearly defined it is, the more likely

³Note that while Vallduví & Vilkuňa (1998) state that rhemes *can* contain contrasts, they do not claim that they have to.

the kontrast is intended to be contrastive. A contrastive reading may be more available if at least one of the alternatives has been explicitly mentioned (as with *2001* above); or, importantly, by manipulating prosodic prominence (e.g. *movie* above). There is no essential difference in the semantics of an utterance where the alternatives to the kontrast are clearly definable, and one where they are not. Rather, all kontrasts introduce a set of alternatives, but the extent to which these are resolved and enumerated varies (cf. Gundel & Fretheim 2004, p.190-1).

Alternative sets involve presupposition. That is, the set of alternatives should be relevant to the discourse at that point. If the implied presupposition of alternatives is not relevant, then it must be accommodated, giving rise to conversational implicature (cf. Atlas 2004). We saw this earlier with (1): by making the rhematic kontrast *movie*, the speaker introduced a presupposition of alternatives to *movie*, e.g. *book*, *play*. The discussion to that point had been entirely about *movies*, so these alternatives were not relevant; therefore they needed to be accommodated. This introduced the intended implicature, i.e. why *2001* is not a good movie. This proposal is in the spirit of Gundel & Fretheim's (2004, pp.192-3) suggestions about topic-focus structure (our theme-rheme): "topic-focus structure is exploited at the grammar-pragmatics interface, where information expressed in the proposition is assessed in order to derive contextual effects" (see also Rooth 1992, Roberts 1996). In our analysis below we take this point further, to contend that in manipulating prosodic prominence, speakers can not only affect information structure per se, but also contrastiveness, and therefore its pragmatic effects.

1.2 Prosodic marking of Contrast

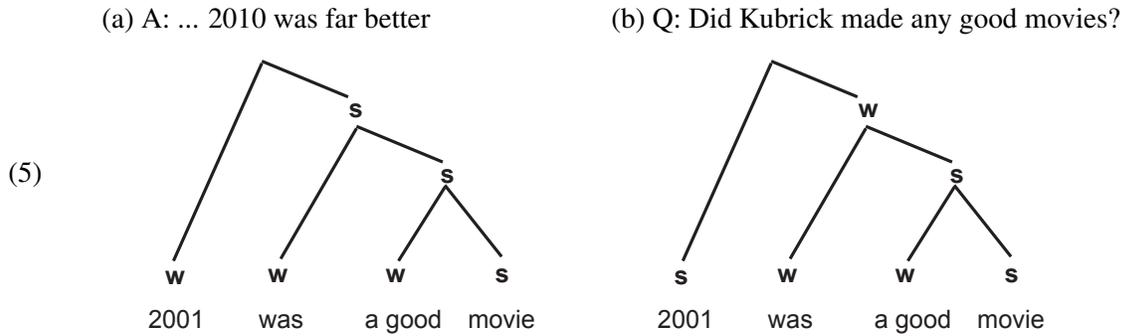
In English, the kontrast in each utterance is standardly held to be marked by a *pitch accent* (e.g. Halliday 1968, Jackendoff 1972, Rooth 1992, Selkirk 1995, Ladd 1996, Steedman 2000). Syllables with pitch accents are *prominent* or stressed at a phrasal level, i.e. they have higher pitch, larger pitch movement, greater duration and intensity than other syllables in the phrase (e.g., see Terken & Hermes 2000). However, we claim that accents do not occur purely for functionally-motivated reasons (the so-called 'focus-to-accent' approach, see discussion in Ladd 1996, ch. 6). Rather, they are manifestations of strong nodes in metrical prosodic structure, and prominence-marking involves manipulation of metrical structure as a whole, not just accent placement (see Ladd 1996, ch.6, Calhoun 2006, 2008, contra Selkirk 1984, Pierrehumbert & Hirschberg 1990). In each phrase, syllables align

with a binary-branching structure of prominences, with the strongest node being the *nuclear* accent (Lieberman 1975). Other strong nodes may also be realised with accents, however in post-nuclear position these tend to have much lower pitch and little or no pitch movement (Grice, Ladd & Arvaniti 2000).

Metrical structure has important properties which affect both the production and perception of prominence (see further in Calhoun 2008, Calhoun 2006, ch.3). In English, the structure is by default *right-branching* (Lieberman 1975, Ladd 1996). This means that speakers are more likely to place the nuclear accent at the end of the phrase, and biases the right-most accent to be heard as the most ‘structurally’ prominent, even if it is not the highest, longest or loudest (Lieberman 1975, Rump & Collier 1996, Ayers 1996). Metrical structure is *rhythmically* constrained: stress is manipulated to maintain an approximately alternating pattern of strong and weak nodes at each level of the structure (see Hayes 1995, Shattuck-Hufnagel, Ostendorf & Ross 1994). These properties constrain how prominence is used functionally, e.g. to mark contrast; and, as we shall see, our expectations about prosodic structure affect how prominence is interpreted, e.g. in determining contrast scope.⁴

Under a metrical account, the basic relationship is between contrast-marking and nuclear prominence, rather than with accenting per se (Ladd 1996, Truckenbrodt 1995, Calhoun 2006, Calhoun 2008). That is, in the alignment of words with metrical structure, contrasts try to align with nuclear positions. For example in (1), *movie* is the rhematic contrast, and aligns with the nuclear accent, which falls in its ‘default’ position on the right-most content word (i.e. (5a), we return to the marking of *2001* below). However, if (1) had instead been uttered in the context in (5b), the only contrast would be on *2001*. This would result in ‘metrical reversal’, i.e. reversal of the usual *weak (w)-strong (s)* ordering at the highest level of structure, so that the nuclear accent would fall on *2001*, as follows:

⁴Note that this approach is opposed to the ‘accent-first’ approach (e.g. Selkirk 1984, Selkirk 1995), where the right-branching bias (or nuclear stress rule), can be overridden by contrast-marking (her focus) when the contrast is not on the last word in the utterance. Our argument is that the right-branching bias applies in all cases, because when the nuclear accent does not fall in the expected position because of contrast-marking, this has interpretative effects. However, it is beyond the scope of this article to provide full arguments for our approach over the ‘accent-first’ approach (see Ladd 1996, Calhoun 2006, Calhoun 2008).



The limit of the rheme (or theme) unit (focus projection) is the nodes which the nuclear accent dominates in the metrical structure (Ladd 1996, Truckenbrodt 1995, Büring 2006). In both (5a) and (5b), since the kontrast is on the nuclear accent, the rheme unit can project to the whole phrase; although in (5a) it does not include *2001* as this is also kontrastive. Kontrast scope, i.e. what elements are intended to be part of the alternative set, is determined dynamically from prominence and contextual cues. In (5a) the prominence structure is expected, because of the right-branching bias. This means the kontrast could have scope over the whole rheme phrase; however, because of the context (e.g. the parallel with *book*), it is probably limited to the word *movie* (cf. (4) where the whole utterance is kontrastive). In the reversal case (5b), the prominence pattern is unexpected, blocking extension of the scope of the kontrast rightward. We develop this idea in the next section and see further examples in section 3.3.

The phonetic prominence of an accent can also vary gradiently. This does not necessarily affect the underlying metrical structure. However, it can affect kontrast status and contrastiveness. For instance, the metrical account implies that pre- and post-nuclear accents cannot signal kontrast. This clears up many problems with ‘optional’ accents for straight ‘focus-to-accent’ accounts (see Calhoun 2008, Calhoun 2006, ch.2). However, the very prominent pre-nuclear accent on *2001* in (1) does mark a thematic kontrast (see Figure 1). This does not affect the metrical structure, (5a), as *movie* is still nuclear even though it has much lower pitch. Phonetic prominence also affects the interpretation of nuclear accents. As the phonetic prominence of a nuclear accent increases, the scope of the kontrast narrows. A particularly high (or emphatic) nuclear accent fairly unambiguously marks ‘narrow focus’, i.e. a kontrast with scope over only the accented word (cf. Eady & Cooper 1986, Rump & Collier 1996, Xu & Xu 2005). In the next section, we will argue that increased prominence not only increases the

likelihood of a kontrast, but also a contrastive interpretation and resulting pragmatic effects. In fact, ‘contrastive’ or ‘emphatic’ accents are often defined as producing contrastive readings. These accents have been described as ToBI L+H*, as opposed to H* (e.g. Pierrehumbert & Hirschberg 1990, Bartels & Kingston 1994, Watson, Gunlogson & Tanenhaus 2008). However, evidence that they really form a separate category, as opposed to being at the extreme of a gradient scale of prominence, is inconclusive (e.g. Ladd & Morton 1997, Kraemer & Swerts 2001). We leave this issue open here.

2 The Probabilistic Interpretation of Prominence

In the discussion above, we outlined our account of how prosody signals kontrast and the possible pragmatic effects of the interaction of the information structure and prosodic systems. Here, we develop this further: we claim that contrastiveness is derived from probabilistic expectations about prominence, given syntactic and contextual cues. Further, speakers can exploit these expectations to manipulate presuppositions in the discourse model and create conversational implicatures.

Our starting point is to regard the mapping between words (or syllables) and metrical structure as probabilistic, with kontrast acting as a strong constraint on the mapping. We can then conceptualise contrastiveness as arising from the *likelihood* of a given word-structure mapping. Broadly, if a word is more prominent than expected given its properties and position in the metrical structure, then it is more likely to be contrastive; conversely, if a word is less prominent than expected, it is more likely to be backgrounded. Expectations are built on the prosodic, syntactic and discourse properties of the words in the utterance. The precise nature and strength of the relevant factors are not yet understood. However, there is evidence for a range of plausible constraints. For instance, in the last section, we laid out constraints on prosodic structure, such as rhythm and the right-branching bias. At the lexical level, certain words seem to be more ‘accentable’ than others: features such as part-of-speech type, word probability (the frequency of the word type divided by the number of words in the corpus) and bigram probability (the likelihood of a word given the word before) are highly predictive of accents (Hirschberg 1993, Pan, McKeown & Hirschberg 2002, Chen & Hasegawa-Johnson 2004). Nenkova, Brenier, Kothari, Calhoun, Whitton, Beaver & Jurafsky (2007) have recently shown that a word’s accent ratio, i.e. the likelihood that its lexeme is accented, is a very good predictor of accenting in

the Switchboard corpus. Givenness relative to the discourse may also affect accentability (Terken & Hirschberg 1994, Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon & Newlands 2000, Baumann & Grice 2006). At the syntactic level, syntactic boundaries tend to align with prosodic boundaries (see review in Shattuck-Hufnagel & Turk 1996); more so for some clause types than others, e.g. adjuncts versus complements (Price, Ostendorf, Shattuck-Hufnagel & Fong 1991). Further, heads of syntactic constituents are less likely to be accented than their arguments, and accents tend to fall at the end of syntactic clauses (Pan et al. 2002).

With these constraints in hand, we can fully develop our theory of firstly how the position and scope of *kontrast* is marked; and then what promotes contrastive readings. Words that are lexically or syntactically “strong” are more likely to be *kontrastive*. For instance, nouns and adjectives are more likely than other words to be both accented and *kontrastive*. This makes sense since arguments more typically have alternative sets. Words that are unpredictable (which can be measured roughly in terms of word or bigram probability) are probably more likely to be *kontrastive* as they contain more information (cf. Bolinger 1972). The properties of syntactic structure act in consort with prosodic structure to serve information structure (cf. Halliday 1968). Words at the end of syntactic clauses are more likely to align with nuclear accents and to be *kontrastive*. By default, nuclear accents fall at the end of prosodic phrases. Therefore, they are likely to fall at the end of syntactic clauses; and on arguments that occur at the end of clauses in English, such as objects.

This confluence between syntax and prosody, however, says nothing about the *scope* of *kontrast*, and therefore the salience of the alternative set. In fact, utterances where our ‘default’ expectations about prosody, syntax and *kontrast* are met are ambiguous as to *kontrast* scope (or ‘broad’ versus ‘narrow’ focus), e.g. *movie* in (1) (though, as we said in the last section, the *kontrast* is likely limited to *movie* because of the parallel co-ordinated syntactic structure, ... *movie*, ... *book*). However, given the context, a *kontrast* on *movie* itself is not expected, as we outlined in section 1.1. Therefore, this makes the alternative set of *movies* salient (despite the fact that the accent on *movies* is not very phonetically prominent), introducing an implicature (that its dependence on the book is the reason A thinks *2001* is not a good movie).

Having built up our expectation-based account, however, we can look at this interaction of prosody, semantics and pragmatics in a different way. We suggest that pragmatic effects of contrastiveness arise

because speakers use probabilistic expectations about the prosody/information structure interface precisely in order to get those pragmatic effects. Speakers indicate through the manipulation of expected prominence that their utterance means more than its propositional content/information structure suggests. This draws attention to the contrastiveness of the *kontrast* and therefore intended implicatures arising from the presupposition of its alternative set. This in turn aids discourse coherence.

Where the *kontrast* falls in an unexpected position in prominence structure, the prosodic, semantic and pragmatic systems are even more intertwined. This usually also means the *kontrast* has ‘narrow’ scope and that the alternative set is salient. In (5b), the metrical reversal leads not only to the *kontrast* being on *2001*, but also makes the alternative set of *movies by Kubrick* salient. This leads to the implicature that his other movies were not very good. Again, it may even be the intended pragmatic effect that is driving the prosody/information structure relationship. Similarly, pre- and post-nuclear accents do not normally signal *kontrast*; however, they can if they are more prominent than expected given their other properties (e.g. the strong accent on *2001* in (1), see Figure 1). Once more, these cases are also contrastive. Finally, even where the *kontrast* falls in an expected position in metrical structure, increased phonetic prominence can be used to force a contrastive reading (as we will see in the next section). Crucially, this increase is not simply any absolute increase in prominence, but is rather relative to the *expected* prominence for the word.

In summary, our aim is to show how contrastiveness is marked in speech, and what the pragmatic effects of contrastiveness are. Contrastiveness arises in the first place from *kontrast*-marking. All *kontrasts* are theoretically contrastive, in that they introduce a presupposition of alternatives. Words marked with nuclear accents are normally *kontrastive*, though pre- and post-nuclear accents can mark *kontrast* where they are more prominent than expected (either in terms of increased phonetic prominence, or because of their lexical/syntactic properties). *Kontrasts* give rise to contrastive readings when their alternative set is particularly salient. The more prominent a word than expected (on the basis of its syntactic, prosodic and discourse properties), the more salient the alternative set. For example, when the nuclear accent does not fall in the expected position (given the right-branching bias), the accented word is more prominent than expected. By manipulating the nature and salience of presuppositions arising from alternative sets, speakers derive pragmatic effects, such as conversational implicatures.

3 Contrast in a Spoken Language Corpus

We now go on to test key predictions of the approach developed above using a spoken language corpus, Switchboard (Godfrey, Holliman & McDaniel 1992). The Switchboard corpus consists of spontaneous telephone conversations between pairs of American English speakers on set topics. This corpus was used because it is a large collection of naturally occurring speech that already had a large range of relevant annotations, including the transcription, syntax/part-of-speech, disfluencies and information status (see Calhoun, Carletta, Brenier, Mayo, Jurafsky, Steedman & Beaver to appear). In the first section below we describe our development of the corpus, which involved operationalising and executing *kontrast* and prosody annotation. The annotated corpus was used to carry out both a quantitative and qualitative analysis of *kontrast*. This allows us to see firstly whether the general features of *kontrastive* words are consistent with our theory; and secondly to examine selected examples, to see how *kontrastive* interpretations and pragmatic effects arise in each case.

3.1 Annotation

Kontrast Our *kontrast* annotation scheme sought a balance between the semantic notion of *kontrast* defined above, and the more intuitively accessible idea of *kontrastiveness* introduced in section 1. Annotators were asked to identify 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”. However, rather than marking *kontrast* directly, annotators used discourse contexts which commonly evoke contrast, i.e. words or NPs were marked as one of the following *kontrast* types (based on Rooth 1992), all other words being *background* (listed in order of frequency). Only ‘contentful’ words were marked (nouns, verbs, adjectives, adverbs, pronouns and demonstratives). False starts, hesitations and filler phrases such as “in fact” or “you know” were excluded from analysis (see further in Calhoun et al. to appear, Calhoun 2006, ch. 5).

- *contrastive*: directly differentiated from a mentioned, semantically related word, e.g. *movie* and *book* in (1);
- *subset*: a member of a more general set mentioned in the context, e.g. *the only bulb that repeats well in my area is the daffodil*;

- *adverbial*: a focus-sensitive adverb such as *only* or *even* is associated with the word, e.g. *the only bulb that repeats well...* ;
- *answer*: completes a question (or open proposition) by the other speaker, e.g. *I bet you what those things are is a **dutch iris***;
- *correction*: corrects or clarifies a previous word or phrase, e.g. *now are you sure they're **hyacinths***;
- *other*: is clearly kontrastive, but does fall into any of the above types, e.g. *it was so good that I had forgotten it was **Christmas Eve***.

145 Switchboard conversations were annotated for kontrast by two post-graduate linguistics students at the University of Edinburgh. They were not specialists in discourse semantics, but were given extensive training. Annotators could listen to the conversation (but could not see an acoustic display). This was necessary to identify kontrasts actually intended by the speaker, rather than all potential kontrasts; given the highly ambiguous nature of spontaneous speech. Agreement between the two annotators was measured on three conversations using the kappa (κ) statistic (Carletta 1996). ‘Blind’ agreement (without consultation) was 0.67, after discussion this rose to 0.85. To our knowledge, this is the only large-scale annotation scheme using such a broad definition of focus/kontrast. Previous schemes have used much more restrictive definitions, e.g. new information or explicit kontrasts (Nakatani, Hirschberg & Grosz 1995, Hedberg & Sosa 2007, Zhang, Hasegawa-Johnson & Levinson 2006) (though see Buráňová, Hajičová & Sgall 2000). It is difficult, therefore, to assess what a ‘good’ level of agreement is. However, these results seem to indicate though this is a hard task, there is a reasonable level of basic accordance.

In all subsequent analyses, the different kontrast types are grouped to give a binary distinction between *kontrasts* and *backgrounds*. Preliminary investigations showed these kontrast types behaved similarly with respect to the features of interest in the corpus.

Prosody Prosodic annotation is a well-established task. We therefore based our scheme on the most commonly used standard today, ToBI (Beckman & Hirschberg 1999). However, we made some changes to concentrate on properties of prosodic structure relevant to signalling kontrast (see section 1.2).

As well as marking the presence of accents, in our scheme one accent in each fluent phrase was marked as *nuclear*. The nuclear accent was defined as the most *structurally* prominent, normally the right-most one, not necessarily the most *phonetically* prominent (see section 1.2 and Ladd 1996). Non-nuclear accents could be *weak* or *full* (although all non-nuclear accents are grouped below). Phrase breaks (ToBI break index levels 3 and 4) were used to group words into a single category of prosodic ‘phrase’. Preliminary analysis suggested the two levels of phrasing did not behave consistently enough with respect to the discourse properties being investigated for this distinction to be useful. Words in disfluent phrases (ToBI 1p and 2p), as well as short phrases containing only discourse fillers (e.g. “um” or “you know”) (ToBI X), were excluded from the analysis.

31 Switchboard conversations were annotated for prosody. Most annotations were done by a post-graduate linguistics student at the University of Edinburgh with experience using ToBI, and a small number (3) by the author. One conversation side was used to check agreement between the annotator and the author. Kappa agreement on accent type was 0.8 and on break type 0.89; both levels of agreement showing “good reliability” (Carletta 1996). These scores are commensurate with those reported for previous ToBI annotations (Pitrelli, Beckman & Hirschberg 1994, Yoon, Chavarría, Cole & Hasegawa-Johnson 2004), suggesting the changes made in our scheme were successful.

3.2 Kontrast Modelling

Above we developed our claim that kontrast and contrastiveness are signalled by manipulating the prominence of a word, relative to its expected prominence given its syntactic and discourse features; leading to pragmatic effects. In order for this analysis to hold, we should show that statistical properties of kontrastive words in our corpus are in line with the predictions of our theory. In this section, therefore, we report kontrast prediction models built using our corpus. Our annotation did not encode the salience of the alternative set of each kontrast. Therefore, we cannot model contrastiveness directly. However, since our annotators marked words they believed were intended to have salient alternatives, this is plausibly also a model of words which were contrastive.

A number of studies have built kontrast predictors using this corpus (Calhoun 2007, Nenkova & Jurafsky 2007, Sridhar, Nenkova, Narayanan & Jurafsky 2008, Badino & Clark 2008). These studies were trying to model kontrast as a useful feature in computational speech applications, and hence aim

to maximise prediction accuracy using features which can be derived automatically. For instance, Nenkova & Jurafsky (2007) report accuracy of 76.9%, and show that part-of-speech and acoustic prominence are the most useful features, although manually annotated accents are crucial to make the task computationally tractable. Prior to the availability of this corpus, Zhang et al. (2006) automatically detected ‘focus kernels’ (the novel part of utterance) and ‘contrasts’ (syntactically parallel explicit contrasts) in limited domain dialogues. They found that prosodic prominence, semantic dissimilarity and part-of-speech type were all important features; however, their corpus and definitions are much more restricted than ours. Other studies have predicted prominence patterns for different focus structures in constructed dialogues (e.g. Batliner, Oppenrieder, Nöth & Stallwitz 1991, Wagner 1999); thereby focusing on the effect of prominence variation, not the interaction of lexical features and prominence.

Method Logistic regression models were built to predict the contrast status (*kontrast* or *background*) of words in the corpus. The data set consisted of 18 Switchboard conversations. Only words annotated for contrast in fluent phrases were included, a total of 9289 words from 33 speakers, around 2 hours of speech. Clauses without at least one annotated contrast were excluded. A variety of features which were predicted to be useful by our theory, and which had previously been shown to be useful for contrast prediction, were included.

Syntactic features included *clause type* (main, complement, adverbial, relative, parenthetical) and *constituent type* (subject, predicate, object, adjunct); *position in the clause/constituent* (relative to the number of words in the clause/constituent); whether the word was the *head of its constituent*; as well as broad *part-of-speech* type, which were extracted from the Penn Treebank annotation of the corpus (see further Calhoun 2006, ch.5). We also tested lexical predictability features: *word probability* ($\log(p_w)$) and *bigram probability* ($\log(p_{w_i}|p_{w_{i-1}})$) (measured over the whole Switchboard corpus); and *accent ratio*, i.e. the probability of each lexeme being accented (over 60 Switchboard conversations with accent annotation, Ostendorf, Shafran, Shattuck-Hufnagel, Carmichael & Byrne 2001), where this is significantly different to chance (see further Calhoun 2009). For NPs, *information status*, i.e. *old*, *mediated* or *new* in the discourse, was also used (see Calhoun et al. to appear). Prosodic features included *accent status* (*nuclear*, *non-nuclear*, *none*), and position in the prosodic phrase (in syllables, relative to the number of syllables in the phrase). Finally, we included a measure of acoustic *prominence*. This was

calculated by adding together four main acoustic correlates of prominence: mean f_0 , f_0 inter-quartile range, mean intensity and duration (as we were interested in the overall effect of phonetic prominence, this proved to be more consistent than including each of these correlates separately). Fundamental frequency (f_0) was normalised as a percentage of the speaker's logged range, with outliers excluded. Intensity was relative to the speaker mean, and duration relative to the syllables in the word.

All continuous predictors were Z-transformed. This made their regression coefficients easier to interpret, as they were then all on the same scale, with an equivalent range to our categorial predictors (note that each acoustic correlate of prominence was Z-transformed, so they would make an equivalent contribution to the *prominence* measure, then the *prominence* measure was itself Z-transformed). We also tested for collinearity between our features, as high correlations between features in regression models can make regression coefficients unstable and lead to spurious findings of significance (e.g. Hocking 2003). Using the Variance Inflation Factor (VIF) and Tolerance metrics, the level of collinearity between our features was acceptable, i.e. VIF was less than 4 for all features (Myers 1990), and Tolerance greater than 0.2 (Menard 1995).

Results and Discussion Logistic regression models were built using the features just described. First each feature was tested, then each possible two-way interaction between features. Only significant features are included in the reported model. The log likelihood measure is used to show model effectiveness (i.e. how much variation in contrast status it captures). The final model had significantly lower log likelihood than a baseline model (all *background*), $LL = 8350, \chi^2(29, N = 9289) = 4205, p < 0.0005$. This model had a classification accuracy of 79.0% (74.0% of contrasts correctly classified, and 82.4% of backgrounds); higher than other reported results, although our data set was more restricted.

Table 1 shows the change in the log likelihood of the final model with the removal of each of our significant features (with replacement). The greater the change, the more important that feature in the model. Table 2 shows the direction and magnitude of the effect of each feature in the final model on the likelihood of a contrast. For each feature, the odds ratio (*OR*) is reported, showing the weight of that feature in the model, equivalent to β -coefficients in linear regression. For ease of interpretation it is converted to *% diff*, i.e. the percentage difference in the likelihood of a contrast with that variable. For categorial features (in *sans serif*), this is the effect of the value of the variable given, e.g. if the

Feature	Change in LL	d.f.
word probability	295.2	1
accent status	230.0	2
position in clause	150.4	1
accent ratio	124.9	1
prominence	115.7	1
constituent head status	96.2	1
part-of-speech	93.5	5
accent status <i>by</i> part-of-speech	58.7	10
information status	55.9	3
constituent head status <i>by</i> position in phrase	24.6	1
information status <i>by</i> position in phrase	22.7	3

Table 1: The change in log likelihood (LL) with the removal of each feature in the full kontrast regression model, along with the degrees of freedom (d.f.) for each feature. The bigger the change the more important the feature in the model. All changes are highly significant ($p < 0.0005$).

value of the *part-of-speech* feature is *noun* the likelihood of a kontrast increases 8.8% (compared to the average over all parts-of-speech, only significant levels of categorical variables are shown). For continuous variables, this is a one unit increase on a Z-transformed scale (i.e one standard deviation for that variable).

As we claimed in section 2, a word is more likely to be kontrastive if it is lexically or syntactically ‘strong’. The most important feature, in terms of model fit, is *word probability* (Table 1). As Table 2 shows, the more frequent a word, the less likely it is to be kontrastive, although the magnitude of this effect is moderate (12.6%). *Accent ratio* was also fairly important to model fit (Table 1). The more likely the lexeme of a word is to be accented, the more likely also that it is kontrastive (Table 1). *Part-of-speech* type had a smaller effect on model fit, though this worked in the way expected. If a word is a *noun* or an *adjective* it is more likely to be kontrastive, as nouns and adjectives are more likely to have alternatives (see section 2); whereas if it is an *adverb* it is less likely to be kontrastive. In line with the discussion in section 2, *position in clause* and, to a lesser extent, *constituent head status* were both important features in the model (Table 1). Kontrast likelihood increases the later the word in the clause, and is lower for constituent heads (Table 2). *Information status* has a small but significant effect on model fit. As expected, if a word is *new* it is more likely to be kontrastive, if it is *old* it is less likely (recall this annotation only applied to nouns). Kontrast likelihood is also higher for *mediated* words,

	Feature	OR	OR:CI	% diff	Wald
Increase	nuclear accent	3.48	2.96-4.09	28.5%	228.6**
	nuclear accent <i>by</i> verb	2.17	1.56-3.01	18.8%	21.5**
	non-nuclear accent <i>by</i> pronoun	2.12	1.16-3.88	18.3%	5.9*
	adjective	1.90	1.51-2.38	15.7%	30.2**
	non-nuclear accent	1.74	1.49-2.03	13.7%	49.3**
	nuclear accent <i>by</i> adverb	1.64	1.08-2.49	12.3%	5.4*
	non-nuclear accent <i>by</i> verb	1.60	1.15-2.22	11.7%	7.9**
	new	1.54	1.33-1.79	10.8%	32.8**
	accent ratio	1.46	1.37-1.57	9.5%	120.3**
	position in clause	1.46	1.37-1.54	9.3%	148.9**
	prominence	1.38	1.30-1.46	8.0%	112.4**
	noun	1.37	1.12-1.68	7.9%	9.5**
	constituent head <i>by</i> position in phrase	1.35	1.20-1.52	7.5%	24.6**
	mediated	1.31	1.16-1.48	6.7%	19.1**
new <i>by</i> position in phrase	1.26	1.10-1.44	5.7%	11.0**	
Decrease	adverb	0.46	0.34-0.61	-18.2%	28.3**
	constituent head	0.55	0.48-0.62	-14.3%	95.9**
	non-nuclear accent <i>by</i> noun	0.59	0.40-0.87	-12.7%	7.0**
	word probability	0.59	0.55-0.63	-12.6%	230.0**
	old	0.72	0.61-0.84	-8.1%	16.9**
	mediated <i>by</i> position in phrase	0.80	0.71-0.88	-5.7%	17.9**
	Constant	0.72	-	-	64.8**

Table 2: Direction and magnitude of the effect of each feature in the contrast regression model. The odds ratio (OR), confidence interval of the odds ratio (OR:CI) and the percentage change in likelihood (% diff) are given, along with the Wald statistic (a measure of the significance of each feature: ** = $p < 0.01$, * = $0.01 \leq p < 0.05$). For categorical features the effect of (significant) values of that feature are given (in sans serif). For continuous features the effect of a one unit increase is shown (on a Z-transformed scale).

i.e. words that had not been mentioned before, but could be inferred from the previous discourse or general knowledge. This may be because its common sub-types, e.g. *set* or *situation* (see Calhoun et al. to appear), are often contrastive as they pick out the part of a set being discussed. Another common subtype, *general*, covers commonly known entities that may nevertheless behave similarly to *new* entities. The interaction between information status and information subtype was not significant, though this may be due to data sparsity in some subtypes.

As expected given the discussion in section 1.2, *accent status* is the second most important feature in the model (Table 1). If a word has a nuclear accent then it is substantially more likely to be contrastive (28.5%) (Table 2), it is also more likely to be contrastive if it carries a non-nuclear accent

(13.7%). Acoustic *prominence* also significantly affects model fit and the likelihood of a kontrast, however the effect of acoustic prominence is smaller in both importance and magnitude than metrical accent status.⁵ This consolidates our claim that the perception of the prominence is primarily through its place in prosodic structure, above and beyond the actual phonetic prominence. However, the likelihood of a kontrast is not solely determined by its prominence (both structural and phonetic). In line with our claim that expectation plays a role in how prominence is interpreted, we find that the interaction *accent status by part-of-speech* has a significant effect on model fit (Table 1). Generally pronouns are unaccented and backgrounded. Therefore, if accented, a kontrast is much more likely (*non-nuclear accent by pronoun*, Table 2). Likewise, while *adverbs* are in general unlikely to be kontrastive, if they carry a nuclear accent this likelihood increases considerably (*nuclear accent by adverb*). Similarly, most verbs are unaccented (53.7% in this data set), so if a verb has a nuclear or non-nuclear accent then a kontrast is more likely (*nuclear accent by verb* and *non-nuclear accent by verb*). On the other hand, since most nouns carry nuclear accents (53.1% in this data set), if a noun has a non-nuclear accent then a kontrast is actually less likely (*non-nuclear accent by noun*).

Although it is intimately linked to prominence perception (see section 1.2), there is no independent effect of phrase position on kontrast likelihood. However, we do see interesting interactions between *position in phrase* and *constituent head status*, and *position in phrase* and *information status* (Table 1). While overall a constituent head is less likely to be a kontrast, this increases considerably toward the end of the phrase (*constituent head by position in phrase*, Table 2). The head is then in an unexpectedly prominent position, and is therefore more likely to be kontrastive. The likelihood for new entities also increases even further towards the end of the phrase (*new by position in phrase*), entering the usual position of kontrast. Conversely, the likelihood for *mediated* words decreases towards the end of the phrase. This slightly unexpected result is probably due to lack of homogeneity within the mediated class (see above).

⁵Note that we also tested the combined model with the acoustic *prominence* measure only, omitting all prosodic structure features (accent and phrase position). This led to a significantly less accurate model.

3.3 Illustrative Examples

In this section, we discuss extracts from the Switchboard corpus, to illustrate how speakers signal that they intend a word to be contrastive, and the pragmatic effects of this marking. Using these examples, we can not only show how our theory works in naturally-occurring speech, but also discuss the phonetic manifestation of the prosodic constructs we have introduced above. The examples are from a conversation between A, a 28 year-old man from the Northern US, and B, a 51 year-old woman from the Western US. The general topic is the US federal budget, which at the time (c. 1991), was in deficit.

The first extract shows how the effect of prominence on a word depends on its lexical and syntactic properties and the prominence of the words around it. A and B are discussing the education budget. B is arguing that people should be educated on the job, rather than at college, and A is countering that this would be too expensive for many businesses:

- (6) A: One of the biggest things now is paralegals... trying to get more people in that field. But they can't just bring somebody in, without even having been to school in that area...
- B1: Excuse me I see it being done...
- B2: **I know a friend that works for my lawyer,**
- B3: that has no training whatsoever,
- B4: **and she's training her**

Figure 2 shows the acoustic representation (sound file *lawyer*), along with the accent and phrasing, and kontrast annotation (see section 3.1).⁶ We will discuss two utterances: B2 and B4. B2 has two kontrasts, the thematic *friend* and rhematic *lawyer*. *friend* is a noun, but also a head (of the NP *a friend ... lawyer*). However, it is also 'new' and a head at the end of a phrase (and therefore nuclear). Although the word *friend* is fairly frequent (i.e. high word probability) it also has a high accent ratio (92% of all tokens of *friend* in the corpus are accented). This makes it, on balance, likely to be kontrastive according to the factors in our prediction model (see Table 2). The speaker still, however, realises *friend* with a high F0 peak, i.e. probably more prominence than needed to mark the kontrast. This

⁶Recall we do not distinguish between levels of prosodic phrasing, see section 3.1. Note also that the *f0* scale used in each of our examples is different because they are said by different speakers with different *f0* ranges. All annotation was done independently of the author following the guidelines set out in section 3.1.

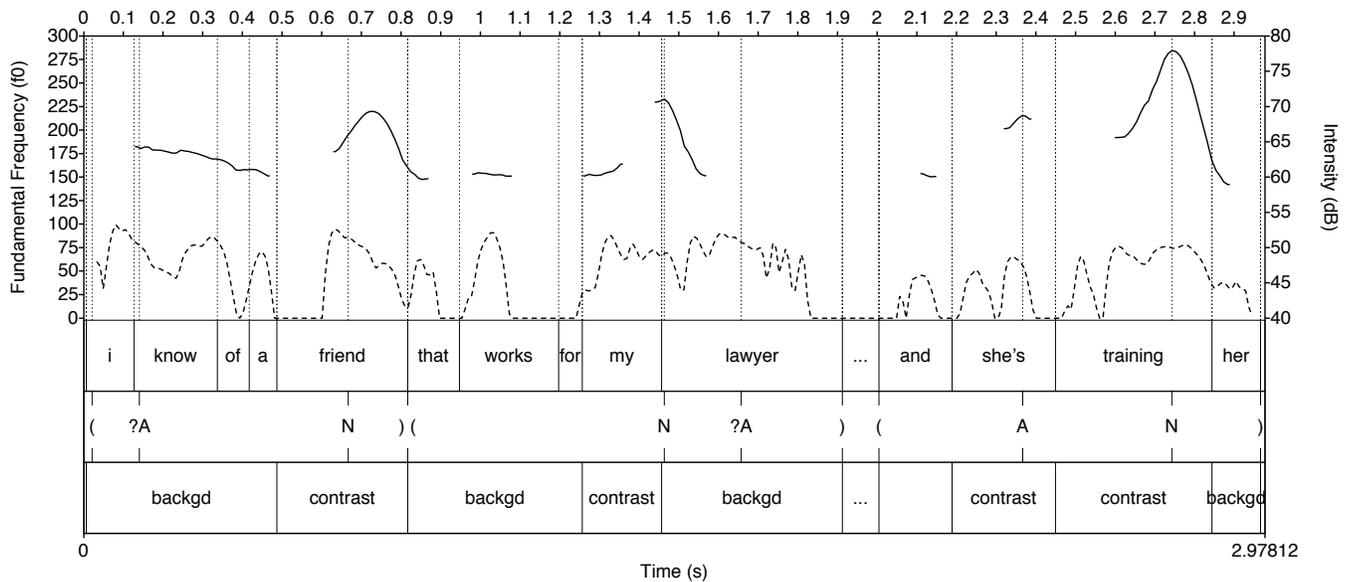
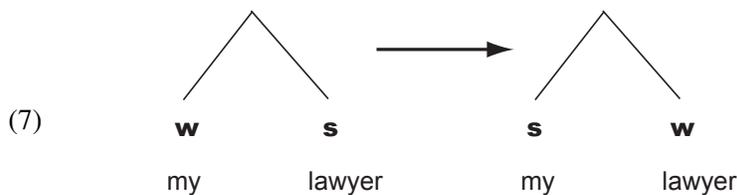


Figure 2: f_0 trace (solid line) and intensity curve (dashed line) for (6), along with the word transcript, accent type and phrasing, and contrast type annotation.

makes the contrast with *they* salient, i.e. companies that are not willing to bring in employees without college education. In the second phrase, the nuclear accent would normally fall on *lawyer*. However, since *lawyer* is inferable (from *paralegals*), metrical prominence is reversed and the nuclear accent falls on *my*:



Note that though *my* is nuclear, *lawyer* is still annotated as weakly accented (shown by its relatively high intensity). It is very unexpected for *my* to have a nuclear accent, since it is a pronoun and has a low accent ratio (21% of tokens of *my* are accented) (cf. Table 2). This not only marks *my* as a contrast, but makes its alternatives salient, i.e. *your [A's] lawyer/lawyers you [A] know*; as we argued in section 2. However, while the reversal marks *lawyer* as backgrounded, its weak accent does not affect its information status. We expect a noun like *lawyer* to be accented (see section 2), especially at the end of a clause and a phrase. Therefore, it is still much less prominent than expected given its

properties.

In B4 the rhematic kontrast is *training*. It is nuclear, as expected, since *her* is pronominal. As a nuclear accented verb it is likely to be kontrastive (cf. Table 2). Again, though, it also has a high F0 peak. This makes its alternative set salient, i.e. as opposed to *not training her*. The phrase has another (theme) kontrast, on *she* (as marked by the annotator). *She* is accented, unexpected for a pronoun, making it likely to be both kontrastive (cf. Table 2) and contrastive (cf. section 2). Once more, we see that it is not just the prosodic prominence of a word per se that determines if it will be contrastive, but also its prominence relative to that expected for a word with its lexical, syntactic and discourse properties.

Why does the speaker make the alternative sets of *friend*, *lawyer* and *training* especially salient? The ‘emphatic’ accents are not necessary to signal the intended information structure, as we have just shown. Nor do the implied alternatives seem to involve any change in topic or resultant disruption in discourse coherence (cf. *movie* in (1)). We contend that B is exploiting expectations at the information structure/prosody interface to emphasise the presupposed alternatives and thereby derive pragmatic effects. She may wish to imply that A is wrong, ignorant or possibly elitist to believe college education is necessary to become a paralegal. This implication arises from the (not strictly necessary) highlighting of the contrasts between: B’s friend and the people A knows, B’s lawyer and A’s lawyers, and *training* as opposed to formal education. B is trying to emphasise how different the alternatives within each of these sets are and so make her implicature. The contrast on *she* may further imply that women lawyers give this kind of opportunity to other women, while male lawyers (and A) do not.

Our second example considers the effects that can come from a salient alternative set when the theme or rheme unit spans multiple phrases. (8) occurs shortly after (6) when B changes the subject from education to the budget deficit, which the speakers had been discussing earlier.

(8) B: you threw that question on me about the deficit... what would you do?

A1: my perception of the budget... the government... has so much money to spend, and there’s not enough money to spread around but

A2: **the deficit basically is the trade surplus between the other countries**

Figure 3 shows the acoustic representation, with annotation as before (sound file *surplus*). There

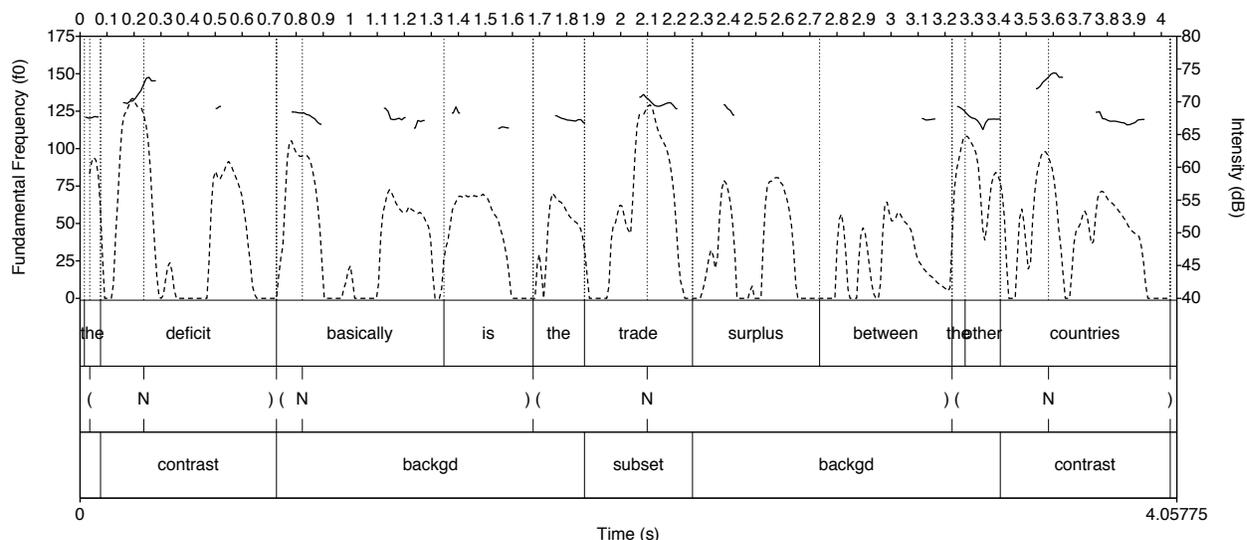


Figure 3: f_0 trace (solid line) and intensity curve (dashed line) for (8), along with the word transcript, accent type and phrasing, contrast type annotation.

are two information units: the thematic *the deficit* and rhematic *the trade ... countries* (we analyse the adverbial *basically* as a hedge that is not part of the propositional meaning). *Deficit* is in its own phrase, fairly unusual for a subject.⁷ It is thereby a noun in nuclear position marked with a high accent, making it likely to be a kontrast (cf. Table 2).⁸ The increased prominence also increases the salience of its alternative set, i.e. contrasting *the budget* with *the deficit*. The pragmatic effect of this is primarily to indicate that the speaker wants to change the topic. However, there may be another intended implicature: that his upcoming assertion is more relevant to the general topic, i.e. addressing the budget deficit, than hers; that is looking at economic causes rather than budget cuts (which was the only solution discussed to this point).

The rheme unit comprises two prosodic phrases. *countries* is clearly likely to be a kontrast, as it is a noun that is phrase final and clause final, and marked by a very high nuclear accent (cf. Table 2). Moreover, the extra prominence (marked primarily by high pitch) makes it contrastive. However, we contend that *trade* is also intended to be kontrastive. It is unexpectedly prominent for a head noun, and

⁷In our data set (see section 3.2), 9.6% of subjects are followed by a phrase break, as opposed to 67.4% of objects, 62.9% of adjuncts and 24.5% of predicates.

⁸Note that the accent ratio of all three key nouns in this example, *deficit*, *trade* and *countries*, is 0.5, and their word probability is only slightly below the average over all words in our data set. Therefore we do not think that accent ratio and word frequency play a big role in this example.

though its accent is not as high as the one on *countries*, it is the loudest in the phrase, and fairly long; making it equally prominent (in the opinion of the author). The salient alternatives therefore include both contrasts with *trade* and *other countries*, as follows:⁹

- (9) [the deficit is [the trade surplus between the other countries]_K] ≡
 is(deficit,trade surplus between the other countries) ⊃
 ¬[is(deficit,fiscal surplus between the other countries) |
 is(deficit,trade surplus in the US) | ...]

The speaker is emphasising that both aspects of his assertion, *the trade surplus* and *other countries* are important to his point. Once more, the intended implicature is probably to contrast these properties with B's assumption that the deficit is best addressed by budget cuts. By viewing the signalling of contrast and contrastiveness as a matter of manipulating expectations about prominence and syntactic/discourse properties, we can get a more nuanced interpretation than standard analyses which assume a direct relationship between prominence and contrast, and therefore a single point of contrast with variable scope.

4 Conclusion

We have looked at what makes a word contrastive from prosodic, semantic and pragmatic perspectives. We claimed that all contrasts are theoretically contrastive, but that they differ in how salient the members of their alternative sets are. The more prominent a word than expected (on the basis of its syntactic/discourse properties), the more salient the alternative set, and therefore the more likely a contrastive reading. Prominence is perceived in relation to metrical prosodic structure, rather than directly from phonetic cues. We further argued that speakers exploit probabilistic expectations about the prosody/information structure interface in order to derive pragmatic effects. They do this by manipulating the nature and salience of presuppositions arising from each alternative set; giving rise to conversational implicatures.

In the second part of the paper, we investigated contrast in a spoken language corpus, Switchboard.

⁹For ease of exposition we represent *trade surplus with other countries* as a single argument, though in a full derivation this would need to be broken up into separate, or nested, predications.

We added new layers of annotation to the corpus, for contrast and prosody. We then presented results of a contrast prediction model showing that a word is more likely to be contrastive if it is more prominent, in terms of its structural position and acoustic prominence, than expected given its syntactic/discourse properties. Finally, we analysed two extracts from the corpus, showing once more how the actual prominence of a word interacts with its expected prominence to derive a contrastive interpretation and evoke conversational implicature.

We hope that our approach provides a useful framework for investigating the relationship between prosodic, semantic and pragmatic properties of language using spontaneous speech. In particular, we hope to have shown the importance of metrical structure in explaining how prosodic prominence is used and perceived, and how expectations built out of statistical properties of language affect interpretation. While our contrast prediction models showed that the prosodic, syntactic and discourse features affected the likelihood of a contrast in the way that we claimed, in future work it would be good to try to build this into a full computational model of contrast interpretation (cf. Jurafsky 2003): in particular to try to operationalise the effect of context. Further, in our qualitative analysis we showed how contrast scope, and the salience of the alternative set, are affected by expected prominence and context. However, these ideas need to be tested more fully, probably using alternative methodology, such as perception experiments. Finally, in future work, it would be good to see if our probabilistic constraint-based approach works for other languages.

References

- Allerton, D. (1978), 'The notion of givenness, and its relations to presupposition and to theme', *Lingua* **44**, 133–168.
- Atlas, J. D. (2004), Presupposition, in L. Horn & G. Ward, eds, 'The Handbook of Pragmatics', Blackwell, Oxford, UK, pp. 29–52.
- Ayers, G. (1996), Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments, PhD thesis, Ohio State University, Columbus, OH.
- Badino, L. & Clark, R. A. (2008), Automatic labeling of contrastive word pairs from spontaneous spoken English, in '2008 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.

- Bartels, C. & Kingston, J. (1994), Salient pitch cues in the perception of contrastive focus, *in* R. van der Sandt, ed., 'Focus and natural language processing', Vol. 1, IBM Working Papers, Heidelberg, pp. 1–10.
- Batliner, A., Oppenrieder, W., Nöth, E. & Stallwitz, G. (1991), The intonational marking of focus structure: Wishful thinking or hard fact?, *in* 'Proceedings of ICPhS', pp. 278–281.
- Baumann, S. & Grice, M. (2006), 'The intonation of accessibility', *Journal of Pragmatics* **38**(10), 1636–1657. Special Issue 'Prosody and Pragmatics'.
- Beckman, M. & Hirschberg, J. (1999), 'The ToBI annotation conventions', http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.
- Boersma, P. & Weenink, D. (2008), 'Praat: doing phonetics by computer', <http://www.praat.org>.
- Bolinger, D. (1965), Pitch accent and sentence rhythm, *in* I. Abe & T. Kanekiyo, eds, 'Forms of English: Accent, Morpheme and Order', Harvard University Press, Cambridge, MA, pp. 139–180.
- Bolinger, D. (1972), 'Accent is predictable (if you're a mind reader)', *Language* **49**, 633–644.
- Buráňová, E., Hajičová, E. & Sgall, P. (2000), Tagging of Very Large Corpora: Topic-Focus Articulation, *in* 'Proceedings of COLING Conference', Saarbrücken, Germany, pp. 278–284.
- Büring, D. (2006), Focus projection and default prominence, *in* V. Molnár & S. Winkler, eds, 'The Architecture of Focus', Mouton De Gruyter, Berlin/New York.
- Calhoun, S. (2006), Information Structure and the Prosodic Structure of English: a Probabilistic Relationship, PhD thesis, University of Edinburgh. http://www.psy.ed.ac.uk/people/scalhoun/thesis/calhoun_thesis.pdf.
- Calhoun, S. (2007), Predicting focus through prominence structure, *in* 'Proceedings of Interspeech', Antwerp, Belgium.
- Calhoun, S. (2008), The centrality of metrical structure in signalling information structure: a probabilistic perspective. Submitted.
- Calhoun, S. (2009), How does informativeness affect prosodic prominence? Submitted.
- Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M. & Beaver, D. (to appear), 'Developing the Switchboard Corpus in NXT: A rich resource for investigating the semantics, pragmatics and prosody of dialogue', *Language Resources and Evaluation Journal*.
- Carletta, J. (1996), 'Assessing agreement on classification tasks: the kappa statistic', *Computational Linguistics* **22**(2), 249–254.
- Chen, K. & Hasegawa-Johnson, M. (2004), An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model, *in* 'Proceedings of ICASSP', pp. 509–512.
- Eady, S. & Cooper, W. (1986), 'Speech intonation and focus location in matched statements and questions', *Journal of the Acoustical Society of America* **80**(2), 402–415.
- Godfrey, J., Holliman, E. & McDaniel, J. (1992), SWITCHBOARD: Telephone speech corpus for research and development, *in* 'Proceedings of ICASSP-92', pp. 517–520.
- Grice, M., Ladd, D. R. & Arvaniti, A. (2000), 'On the place of phrase accents in intonational phonology', *Phonology* **17**(2), 143–185.
- Gundel, J. & Fretheim, T. (2004), Topic and focus, *in* L. Horn & G. Ward, eds, 'The Handbook of Pragmatics', Blackwell, Oxford, UK, pp. 175–196.

- Gussenhoven, C. (1983), 'Focus, mode and the nucleus', *Journal of Linguistics* **19**, 377–417.
- Halliday, M. (1968), 'Notes on transitivity and theme in English: Part 3', *Journal of Linguistics* **4**, 179–215.
- Hayes, B. (1995), *Metrical Stress Theory*, University of Chicago Press, Chicago, IL.
- Hedberg, N. & Sosa, J. M. (2007), The prosody of topic and focus in spontaneous English dialogue, in C. Lee, M. Gordon & D. Büring, eds, 'Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation', Springer, Dordrecht, pp. 101–120.
- Hirschberg, J. (1993), 'Pitch accent in context: Predicting intonational prominence from text', *Artificial Intelligence* **63**, 305–340.
- Hocking, R. R. (2003), *Methods and applications of linear models: Regression and the analysis of variance*, second edn, Wiley, New York.
- Jackendoff, R. S. (1972), *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge, MA.
- Jurafsky, D. (2003), Probabilistic modeling in psycholinguistics: Linguistic comprehension and production, in H. Bod, J. Hay & S. Jannedy, eds, 'Probabilistic Linguistics', MIT Press, Cambridge, MA.
- Krahmer, E. & Swerts, M. (2001), 'On the alleged existence of contrastive accents', *Speech Communication* **34**(4), 391–405.
- Kruijff-Korbayová, I. & Steedman, M. (2003), 'Discourse and information structure', *Journal of Logic, Language and Information* **12**, 249–259.
- Ladd, D. R. (1980), *The structure of intonational meaning: evidence from English*, Indiana University Press, Bloomington.
- Ladd, D. R. (1996), *Intonational Phonology*, Cambridge University Press, Cambridge, UK.
- Ladd, D. R. & Morton, R. (1997), 'The perception of intonational emphasis: continuous or categorical?', *Journal of Phonetics* **25**, 313–342.
- Lambrecht, K. (1994), *Information Structure and Sentence Form*, Cambridge University Press, Cambridge, UK.
- Liberman, M. (1975), The intonational system of English, PhD thesis, MIT Linguistics, Cambridge, MA.
- Menard, S. (1995), *Applied logistic regression analysis*, Sage university paper series on quantitative applications in the social sciences, 07-106, Sage, Thousand Oaks, CA.
- Myers, R. (1990), *Classical and modern regression with applications*, second edn, Duxbury, Boston, MA.
- 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', Stanford, CA, pp. 106–112.
- 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.

- 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', Red Bank, NJ, pp. 119–121.
- Pan, S., McKeown, K. & Hirschberg, J. (2002), 'Exploring features from natural language generation for prosody modeling', *Computer Speech and Language* **16**, 457–490.
- 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', MIT Press, Cambridge, MA, pp. 271–311.
- 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.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991), 'The use of prosody in syntactic disambiguation', *Journal of the Acoustical Society of America* **90**, 2956–2970.
- Roberts, C. (1996), 'Information structure: towards an integrated theory of formal pragmatics', *OSU Working Papers in Linguistics: Papers in Semantics* **49**, 91–136.
- Rooth, M. (1992), 'A theory of focus interpretation', *Natural Language Semantics* **1**, 75–116.
- Rump, H. & Collier, R. (1996), 'Focus conditions and the prominence of pitch-accented syllables', *Language and Speech* **39**, 1–17.
- Selkirk, E. (1984), *Phonology and Syntax*, MIT Press, Cambridge, MA.
- Selkirk, E. (1995), Sentence prosody: Intonation, stress and phrasing, in J. Goldsmith, ed., 'The Handbook of Phonological Theory', Blackwell, Cambridge, MA & Oxford, pp. 550–569.
- Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. (1994), 'Stress shift and early pitch accent placement in lexical items in American English', *Journal of Phonetics* **22**, 357–388.
- Shattuck-Hufnagel, S. & Turk, A. E. (1996), 'A prosody tutorial for investigators of auditory sentence processing', *Journal of Psycholinguistic Research* **25**(2), 193–247.
- 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.
- Terken, J. & Hermes, D. (2000), The perception of prosodic prominence, in M. Horne, ed., 'Prosody: Theory and experiment. Studies presented to Gösta Bruce', Dordrecht: Kluwer, pp. 89–127.
- 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.
- Truckenbrodt, H. (1995), Phonological Phrases: Their Relation to Syntax, Focus and Prominence, PhD thesis, MIT.
- Umbach, C. (2004), 'On the notion of contrast in information structure and discourse structure', *Journal of Semantics* **21**, 155–175.
- Vallduví, E. & Vilkuna, M. (1998), 'On rheme and kontrast', *Syntax and Semantics* **29**, 79–108.
- Wagner, P. (1999), The synthesis of German contrastive focus, in 'Proceedings of ICPhS', San Francisco.

- Watson, D., Gunlogson, C. & Tanenhaus, M. (2008), 'Interpreting pitch accents in on-line comprehension: H* vs L+H*', *Cognitive Science* **32**(7), 1232–1244.
- Xu, Y. & Xu, C. (2005), 'Phonetic realization of focus in English declarative intonation', *Journal of Phonetics* **33**, 159–197.
- 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.
- Zhang, T., Hasegawa-Johnson, M. & Levinson, S. (2006), 'Extraction of pragmatic and semantic salience from spontaneous spoken English', *Speech Communication* **48**, 437–462.