THE RELIABILITY OF INTONATION TRANSCRIPTIONS:

IMPLICATIONS FOR SPEECH CORPORA

[Suggested running title]

RELIABILITY OF INTONATION TRANSCRIPTIONS

RICHARD T. CAULDWELL

richard@speechinaction.com

Abstract

Transcriptions of speech which aim to show the speaker's intonation are not sufficiently reliable to warrant the time involved in transcribing and inputting into computer corpora. The transcriptions of the same recording by the same person at different times are likely to differ by at least 15%; transcriptions of the same recording by different people using the same description can differ by up to 40%. The differences between transcriptions of the same data by people using different descriptions are more difficult to quantify, and are the subject of this paper. I compare Discourse Intonation and Systemic Intonation transcriptions of the same data. They differ because of (a) the theoretical standpoints of the two descriptions; (b) the categories and realisations that they recognise; (c) the notation that they use (d) the transcription practices that they adopt. The differences can be as high as 45%, making it doubtful that it is worthwhile investing time and effort in producing such transcriptions.

1 Introduction

There has recently been an emergence of interest in standardising the representation of transcriptions of speech in computer corpora (e.g. Leech, Myers & Thomas 1995). In this paper I question whether it is worth storing intonation transcriptions in computer corpora. I do so because I, working with one description (Discourse Intonation: Brazil 1994, 1997), have problems with transcriptions made using other descriptions: baldly put, they obscure features that I am interested in, and highlight features that are not of interest. Equally, I am aware that users of other descriptions (such as those of Halliday 1967, 1970; O'Connor and Arnold 1971; Crystal 1969) have similar problems using my Discourse Intonation transcriptions.

These problems seem to me to indicate that we have to be cautious about expectations that computer software can translate reliably from one description to another. Roach (1994:12) argues that automatic conversion would be valuable, but Roach and Arnfield (1995:160) note that, in relation to translating a British into an American transcription, a full conversion is not possible. In this paper I explore the extent to which conversions would suffer from problems of reliability. I do so by comparing transcriptions of the same recordings done according to two different descriptions (both British).

The transcriptions concerned were done following the conventions of Discourse Intonation (Brazil 1997) and Systemic Intonation (Halliday 1967, 1970, 1994). This comparison is presented in section 4 below. Prior to presenting this comparison, it is important to understand what the components of a description of intonation are. These are outlined and

explained in section 2 below. In section 3, I introduce the notation necessary to understand the comparison. Henceforth I shall refer to Discourse Intonation and Systemic Intonation by the acronyms DI and SI respectively.

2 Components of a Description

Any description of the systems of intonation of English has four inter-related components: theoretical assumptions, categories/realisations, notation, and transcription practice. The product that one sees, a transcription, is the outcome of recorded data being subjected to these components.

2.1 Theoretical assumptions

The theoretical assumptions are typically expressed by statements about the role of intonation in expressing meanings: these statements relate intonation to one or more of *attitudinal, accentual, grammatical,* and *discoursal* meanings.

SI views intonation as being grammatical in function: 'The contrasts made by intonation in English are clearly not lexical … English intonation contrasts are grammatical'… (Halliday 1967:10). SI treats intonation as having a role to play in a large number of grammatical systems: Halliday (1967:54-61) lists forty of these, they include information structure, information focus and commitment.

DI views intonation as discoursal in function. 'the significance of intonation is related to the function of the utterance as an existentially appropriate contribution to an interactive discourse' (Brazil 1984:46). 'By making a choice in any of the intonation systems ... a speaker makes some kind of assumption about what he/she takes, for present purposes, to be the state of understanding between him/her and a hearer' (Brazil 1997:132). Speakers

4

thus make intonation choices according to their perception of the understandings they share with their hearers: these understandings relate to their shared biographies, and to the purposes of their talk in a particular context. Although syntax and intonation do have a relationship in purpose-driven talk (Brazil, 1995), they are regarded as being separate areas of choice. DI does not have a normative stance regarding the relationship between intonation and grammar.

2.2 Categories and realisations

Arising from the theoretical assumptions comes a set of categories which each description chooses to recognise, and for which the description proposes a corresponding set of prosodic realisations.

SI is concerned with the interaction between grammar and intonation, and posits primary categories of the 'information unit', realised by the 'tone-group' and 'key' (e.g. *contradiction*, *reservation*, *command* cf. Halliday 1994:305) realised by tone. The tone group is regarded as having an 'unmarked' relationship with the clause; and tones are regarded as having unmarked relationships with clause-types: for example, falling tone is the unmarked version for statements and 'wh'-questions.

DI is interested in the speakers' moment-by-moment context referenced choices. It recognises categories of *selections of meaning* realised by *prominences* in *tone-units*, and the basic speech functions of *telling* and *referring* realised by choices of tone - for example a falling tone realises an act of telling.

2.3 Notation and standardisation

The specific details of the notation of both SI and DI are outlined below (cf. section 3 below), it is however worth mentioning one or two general points

about notation at this stage. Notation is both a coding system and a perceptual apparatus. As a coding system, it exists to give a typographic or symbolic representation of the prosodic realisations of the categories that the descriptions recognise. Notations vary in the extent to which they are iconic or keyboard-friendly. The notations of the two descriptions compared below were devised in the days of the typewriter, and the symbols used are generally those that can be managed on a standard typewriter keyboard. With the arrival of the word-processor the notation of DI evolved and later publications such as Brazil (1994) used arrow symbols for tone choice. The notation also acts as a perceptual apparatus for the transcriber. It can be compared to a pair of sunglasses that allows certain wavelengths of light through, and bars others. But unlike a pair of sunglasses, notation excludes far more than it lets through. The notation of SI and DI can be thought of as reducing prosody to the aural equivalent of two-dimensional shadows: the basic shapes can be seen, but much detail is omitted. Transcribers are trained, through a process of standardisation, to be hyper-sensitive to the prosodic realisations recognised by the description, and they are de-sensitised to other prosodic phenomena not regarded as significant. A description may thus predispose transcribers to be deaf to prosodic phenomena that have significance in other descriptions. For example SI's *foot* (cf 3 below) predisposes transcribers to hear every word-accent, whereas DI's adherence to the two-prominence tone unit predisposes transcribers to ignore word-accent unless a prominence occurs on one. Transcription is thus in equal proportion an exercise in selective perception and in selective deafness. Any description, to the extent that it is selective in what it chooses to regard as significant, is necessarily limited in scope. Transcription is an act of interpretation of perceptual phenomena. Because

descriptions differ in their theoretical assumptions and the categories they regard as significant, their transcriptions of the same data will be different. The standardisation process for trainee transcribers requires considerable personal investment in hours of ear-training to develop a perceptual apparatus which matches that of the originators of the description. It is thus difficult for transcribers who are accustomed to one description to adjust to the conventions of another description.

2.4 Categories versus Realisation

Although all descriptions aim to be inclusive in capturing all the prosodic phenomena relevant to the categories they recognise, their different transcription practices vary along a 'category/realisation' cline. That is to say that there are descriptions (of which DI is one) which encourage a 'realisation' focus approach to transcription: whenever a transcriber hears an instance of a fall, a fall will be transcribed, whatever problems this might provide later in interpretation. There are other descriptions (of which SI is one) which encourage a 'category' focus approach: wherever a transcriber hears a fall which coincides with a choice in a grammatical system, that fall will be transcribed, but other "falls" may be ignored if they do not co-occur with a grammatical system which the description recognises.

2.5 Methodology

The recordings discussed below come from Halliday (1970): they consist of scripted speech read aloud by actors, who had been coached in producing the intonation features indicated by the notation. After the recordings were completed, the recordings were compared with the script and some alterations were made to adjust the 'transcription' to make it match what the actors had done.

It is therefore reasonable to treat the 'transcriptions' as reliable. There are

however some mistakes in the transcription (e.g. foot boundaries omitted) which could be due to errors in the publication process. These errors are very few and I make an effort in what follows not to attribute any general trends to these errors. Given the relative complexity of the notation and the number of stages between script-writing and publication, a small number of errors are inevitable. The same applies to my DI transcription.

There were three stages to the DI transcription: two separate auditory transcriptions; and a comparison of the two transcriptions to resolve differences. The DI transcription was produced with reference to the recordings alone. There was no recourse to the SI transcription until all the stages of the DI transcription were complete.

The recordings were transcribed twice in a period spanning two months. I used an auditory method of transcription which involved repeated listening to the recordings section by section using a Sony TCMR1 repeatcorder with earphones. The units were transcribed twice, and the two resulting transcriptions were then compared in order to identify conflicts. These conflicts were then resolved (or allowed to stand as cases of "uncertainty") by recourse to the tape, thus producing a 'considered' transcription. The degree of reliability between my initial transcriptions was an overall 85%, with a low of 82 % (for a 'spontaneous monologue'), and a high of 91% (for poetry).

As an additional measure of reliability I compared my 'considered' transcription with an independent transcription done by a colleague. Our different transcriptions had an overall rate of agreement of 70% prior to setting about producing an agreed version. The version "agreed" between

myself and the colleague was 90% in agreement with my considered version. I am thus fairly confident that the transcriptions arrived at were reasonably reliable for the task of comparison of the two descriptions in action.

3 Notation

Before proceeding to the comparison, it is necessary to introduce, briefly, the notation conventions of each description.

3.1 'Stress', and tone groups and tone groups

Neither SI or DI use the term 'stress': SI uses the term *salience* and DI uses the term *prominence.* For SI, a salience is the first stressed syllable in a foot. The following tone group consists of ten feet. The salient syllables are those which come immediately after the slash (/) symbol. The double slash (//) indicates a tone-group boundary. The SI tone group is made up of one or more feet. There is no theoretical limit to the number of feet in a tone group, but in the data used for the comparison, the maximum is ten:

> // ∧ with / ice / ∧ a / slice of / lemon / ∧ and a / tiny / splash of / <u>so</u>da / please //
> (Halliday 1970:138)

Note that three of the ten feet begin with a 'silent ictus', symbolised by a caret (∧). For this stretch of speech, then, SI has ten feet with seven saliences and three silent ictuses. The underlined syllable is the tonic, the location of the tone. We will return to the topic of tones below (cf. 3.2)

DI indicates the presence of prominences by the use of upper case letters on prominent syllables. In the DI version of the tone group quoted above, there are five prominent syllables in three tone units:

> // with ICE // a SLICE of <u>LE</u>mon // and a TIny splash of <u>SO</u>da please //

The prominent syllables are in upper case, tonic syllables are underlined, the tone unit boundary symbol is the same as SI's – the double slash.

9

Emerging from this brief explanation of the notation, there are differences between the descriptions worth noting, and these are summarised in Table 1.

Table 1 differences between SI and DI transcriptions of the same tone group

	SI	DI
Tone units/groups	1	3
Tones	1	3
'Stresses'	7	5
'Silent ictuses'	3	0

Table 1 shows that for the same stretch of speech, SI has one tone group, that DI has three; that SI has one tone and DI has three; that SI recognises more 'stresses' than DI. SI allows pauses to occur within a tone group as silent ictuses, and DI does not: for DI places tone unit boundaries wherever there is a pause, but boundaries also occur where there are no pauses. For SI saliences (by and large) equate to word-stress, for DI prominence does not equate to word-stress: prominence is a speaker choice.

DI uses the term 'tone <u>unit</u>', whereas SI uses the term 'tone <u>group</u>': where it is clear that I am discussing one description and not the other, I shall use the term appropriate for that description. Where I need to refer to both tone units and tone groups I shall prefer SI's term 'tone group'.

3.2 Tone

SI recognises five simple tones and two compound tones, DI recognises the same five simple tones, but not the compound tones. Table 1 shows how SI's tones compare with their counterparts in DI. Both descriptions recognise that there is a fundamental distinction between rising and falling tones – and that the criterial factor is the final movement of the tonic.

10

Table 2 Inventories of Tones in SI and DI compared

SI Code	SI Label	Prosodic Shape	DI Label	**DI Code**
Tone 1	Fall	↘	Fall	p
Tone 2	Rise	↗	Rise	r+
Tone 3	Low Rise	↗	Level	o
Tone 4	Fall-rise	↘↗	Fall-rise	r
Tone 5	Rise-fall	↗↘	Rise-fall	p+
Tone 1 3	Fall + Rise	↘　　↗	[none]	[none]
Tone 5 3	Rise + Fall	↗　　↘	[none]	[none]

There are three points to make about Table 2:

    (a) SI has two codes for different types of rise:

           tone 2, which ends high

           tone 3, 'which starts low and ends mid low' (Halliday

              1970:10)

    (b) SI is ambivalent about the existence of level tone, some

           instances of tone 3 may be level, others will be rises

    (c) SI has compound tones 1 3 and 5 3, DI does not.

The two versions of the tone group discussed above with symbols for tones inserted look like this:

      SI //…1 ∧ with / ice / ∧ a / slice of / lemon / ∧ and a / tiny / splash of / soda / please //

      DI // r+ with ICE // r+ a SLICE of LEmon // p and a TIny splash of SOda please //

In the SI version, the symbol 1 indicates that there is a falling tone on the underlined syllable. The three dots preceding the number 1 indicate that there is a listing pretonic with rises on some (but not all) of the salient syllables (cf Halliday 1970:15; and a critique in Cauldwell, 1993). In the DI version, the r+ symbols signify rising tones on the underlined syllables in *ice* and *lemon;* the p symbol signifies a falling tone on the underlined syllable in *soda.*

It will already be apparent from this brief introduction to the notation of the two descriptions that there are discrepancies (a) in their treatment of the same stretch of recorded speech (b) and in the number and nature of the categories they recognise. SI and DI have other notational conventions for other categories of intonational phenomena: SI's *secondary tones* (Halliday 1970:14-19) and DI's *key* and *termination* (Brazil 1997:chapter 3), but for the purposes of clarity, I exclude these categories and their associated notation from the comparison.

In the following discussion I use the term "tone" (inside double quotation marks) to refer to occasions in the data when the phonetic criteria for a tone appear, but SI does not transcribe a tone.

4 Data

The recordings analysed were four 'Study Units' from Halliday 1970. The study units were:

> Study unit 35. Spontaneous monologue. (The Central Line)
>
> Study unit 36. Constructed dialogue. (In a restaurant part 1)
>
> Study unit 37. Constructed dialogue. (In a restaurant part 2)
>
> Study unit 38. Poems.

These units were chosen because they represent the full system of notation in action of which Halliday says 'it is the form recommended for use (by teacher and students) in writing out original texts, or analysing passages of English speech…' (1970:119)

I then compared the DI transcription with the SI transcription treatments of the following features of the data:

- "stress" (*prominence* in DI and *salience* in SI)
- the number of tone groups and tone units (tonality)

- the location of tones (tonicity)
- tone choice

5 Findings

The presentation will be punctuated by 'summary statements' (bulleted and in **bold**) of the differences between DI and SI. These statements will make the assumption that what is true for the data under study will also be true for other types of data. As all the SI data appears to have been scripted, the summary statements may not be true for unscripted data. The data does include poetry, and a lecture-style 'spontaneous monologue' so I expect the statements to be accurate for readings of poetry and lecture-style discourse. All statements involving percentages use the SI figure as the percentage base; percentages in the 'summary statements' have been rounded to the nearest 5; thus 78% is rounded up to 80%; 53% is rounded up to 55%.

5.1 Saliences and prominences

Given the different approaches to the phenomenon of "stress" mentioned in section 2.1 above, we would expect SI to show a greater incidence of saliences than DI would of prominences, and this is what in fact happens: SI has 45% of all syllables salient, whereas DI has 30% of all syllables prominent. (Note that the figure for saliences includes silent ictuses) The DI transcription has 10% fewer prominences than SI's total of saliences. There is no significant differences between the four study units in this respect. The issue of the 'silent ictus' does provide a complication however if these are removed from the salience count, the SI percentage drops to 40%. Thus

- **DI has at least 10% fewer prominences than SI has saliences**

5.2 Tone groups and tone units

Table 3 shows the numbers of tone groups that DI and SI assigns to each of

the texts, and the percentage (using the SI count as the base) of agreements on the division into tone groups.

Table 3 Agreements on tonality

Study Unit	SI groups	DI units	No. of agreements	Percentage agreement
35	66	88	35	53%
36	68	77	53	78%
37	67	87	53	79%
38	41	60	23	56%
Totals	242	312	164	68%

The last row in Table 3 shows that DI (with 312) had 70 more tone units than SI (with 242) has tone groups. Thus:

- **a DI transcription is likely to have 30% more tone units than an SI transcription**

Table 3 also shows that the percentage of occasions on which SI and DI agree on the location of tone group boundaries was 68%, thus:

- **on 70% of occasions where SI has a tone group boundary, DI will also have a boundary**

The percentage of agreement varies however with text type from a high of 79% to a low of 53%. Study units 36 & 37 have agreement levels of 78% and 79% – these texts are both 'constructed dialogues' with approximately 50 short(ish) turns. Because DI and SI agree in having tone group boundaries at turn endings, the level of agreement is higher than in the monologic texts. The extent of agreement for the monologic texts, Study Unit 35 ('spontaneous monologue') and Study unit 38 (poetry) is much lower at 53% and 56% respectively. Thus

- **in dialogues with short turns, on 80% of occasions where SI has a tone group boundary, DI will have a boundary**
- **in monologic texts, on 55% of occasions where SI has a tone group boundary, DI will have a boundary**

14

The following is an example of where SI and DI agree on the number of tone units:

SI   // 1+ <u>that's</u> just // -1+ standard e / <u>quip</u>ment //
DI   // p but <u>THAT'S</u> // p just STANDard e<u>QUIP</u>ment//
 (Halliday 1970:127)

Notice that although both descriptions agree in having two tone groups, they do not agree on the precise placement of the boundary. The reasons for this difference require too detailed an explanation which would not be appropriate for the comparison reported here. Cauldwell (1993) develops this point further.

The following example shows DI and SI disagreeing on the location of tone unit boundaries:

SI   // - 1 all the / trains / have / been with / <u>drawn</u> //
DI   // r+ and ALL the <u>TRAINS</u> // p had been with <u>DRAWN</u> //
 (Halliday 1970:127)

In this example, the minus sign preceding the code tone 1 in the SI transcription indicates that there is a secondary tone on the pretonic – SI's 'uneven' pretonic. DI interprets this as a rising tone on *trains*.

5.3 Location of tones

In this subsection, I investigate the extent to which DI and SI agree on the location of tones. In order to make a comparison of tones, it is necessary to count separately the two halves of SI's compound tones – thus boosting the SI tone count by 20 – up to 262 from the tone group count of 242 given in Table 3. This will allow us to say that where DI has a tone occurring on each half of an SI compound tone, this will count as a 'match'.

Table 4 Percentage of SI tonics which co-occur with DI tones

Study Unit	Tonics	Agreement
35	71	93%
36	74	91%
37	72	97%
38	45	100%
Total	262	

Looked at from the point of view of SI, there is a remarkable degree of agreement between the two systems on the location of the tonic. On approximately 95% of occasions, SI and DI agree on the location of tonics. The figure for agreement in the poetry reading is 100%: on all of the 45 occasions where SI transcribes a tone, DI also transcribes a tone. Examples of the agreement of the location of the tonic include:

> SI: // 1+ line would have / been very/ popular //
>
> DI: // p the LINE would have been very POPular //
>
> (Hallliday 1970:127)
>
>
> SI: // 13 Yours was the / sherry //
>
> DI: // p YOURS // r+ was the SHErry //
>
> (Halliday 1970:128)

The 5% of locations where there is no match could be attributable to error on the part of the actors, the publisher, or DI transcriber error.

The percentages for agreements given in Table 4 used the SI total of 262 tones as the percentage base. The picture changes when we use the DI total as the percentage base. Table 5 shows the percentages of agreement using the DI total, 312, as the percentage base.

Table 5 Percentage agreement between SI and DI on tonicity.

Study Unit	Tonics	Agreement
35	89	74%
36	77	87%
37	86	81%
38	60	75%
Totals	312	80%

On 80% of occasions where DI transcribes a tone, SI transcribes a tone. This figure is 15% lower than the proportion of agreements viewed from the SI perspective. For the poetry data, Study Unit 38, the extent of the matching drops from 100% to 75% this is because DI has 60 tones – all those that SI identified plus 15 others.

- **On 95% of occasions where SI has a tonic DI will also have one;**
- **In poetry readings, on all occasions where SI has a tonic DI will also have one**
- **DI will have 16% more tones than SI**
- **DI will have 33% more tones than SI in readings of poetry**

5.4 Tones

Table 6 Numbers of matches of DI tones with SI tones

	Totals	Matches	Non-Matches
Tone 1	114	112	2
Tone 2	25	23	2
Tone 3	23	22	1
Tone 4	51	46	5
Tone 5	9	2	7
Tone 1 3	17	16	1
Tone 5 3	3	0	3
Totals	242	211	31
Percent		87%	13%

Table 6 shows the extent to which DI's judgements of tone choice match those of SI. I counted the following co-occurrences as matches: tone 1 with p tone; tones 2 & 3 with r+ tone; tone 4 with r tone; tone 5 with p + tone; tone 1 3 with p tone followed by r+ tone; tone 5 3 with p+ tone followed by r+ tone. The overall degree of matching, using the SI numbers as the percentage base is 87%. From this table it is safe to conclude that:

- **Where a SI tone occurs, there is a 90% chance that the DI transcription will have a matching tone**

Table 7 DI tones matches

	Totals	Matches	Non-Matches
p	**151**	**127**	**24**
p+	**2**	**2**	**0**
r	**47**	**41**	**6**
r+	**79**	**60**	**19**
o	**25**	**0**	**25**
?	**8**	**0**	**8**
Totals	**312**	**208**	**104**
Percent		**67%**	**33%**

Table 7 shows that when we use DI as the percentage base, the corresponding figures show a drop of 20%. From this perspective, two thirds (67%) of DI tones will match with SI tones, but one third (33%) of DI tones will not have a matching SI tone.

Table 7 shows a number of interesting discrepancies: there are no matches for DI's o tone (level). This finding is against expectations, because although Halliday denied in his early (1967, 1970) work that level tone

exists, a later statement creates the expectation that it will match with tone 3:

> The falling/rising opposition may be neutralized, giving a
> level tone which as it were opts out of the choice. This is
> TONE 3. (Halliday 1994:302).

From the evidence of the 1970 recordings, this is not the case. DI's o tone does not co-occur with SI's tone 3: in 24 of the 25 cases it occurs where SI has a salience, in the remaining case it occurs on a non-salience.

- **DI o tones match with SI saliences, not with tone 3**

The o tone is not the only one to be mismatched with a salience: rising, falling, and fall-rise "tones" are all at various times transcribed as SI saliences. Indeed nearly a half of the 104 mismatches involve these "tone"/salience mismatches.

- **15% of DI tones will be transcribed by SI as saliences**
- **15% of DI r+ tones (rises) will be transcribed by SI as saliences**

6 Discussion

I have demonstrated that descriptions will produce transcriptions which will diverge in some cases quite considerably in their interpretation of the same recorded texts. The differences included:

- **DI will have 30% more tone units than SI**
- **in monologic texts DI and SI will agree on tone unit boundaries in only 55% cases**
- **DI will have 16% more tones than SI**
- **SI transcribes 100% of level "tones" as saliences**
- **SI transcribes 15% rising tones as saliences**

These figures represent a poor level of agreement, particularly when one

19

takes into account the fact that both the SI and DI transcriptions were 'considered' transcriptions, that is they can be regarded as internally reliable. The level of agreement for non-considered transcriptions, i.e. those likely to be done by keyboarders inputting great amounts of speech, is likely to be far lower.

6.1 Theoretical assumptions

The reason for the differences between the respective transcriptions are to be found in their theoretical underpinnings. SI's concern is with the role that intonation plays in grammatical systems, and adopts an approach to transcription which highlights the relationship between intonation and grammar, this approach to transcription is one I termed earlier as 'category-focussed'. A consequence of this approach is that prosodic phenomena which occur, but do not have a predicted relationship with grammar, are ignored.

The differences arise because each description foregrounds some features, and obscures other features of the original recording, according to that description's view of how intonation contributes to meaning.

7 Conclusion

It might be argued that the figures I have given above actually represent an adequate level of agreement between descriptions which simple software computations can handle. The different notations can to a certain extent be made compatible, or interchangeable. For example, as we know that a DI transcription is likely to have 30% more tone units than SI, it is a simple task to reduce the DI figure by 30% and increase the SI figure similarly to get the appropriate totals. This might be an adequate procedure if all one is

only interested in numbers of tone units/groups for a whole text. Similarly with level tones and rising tones: knowing that in an SI transcription a certain percentage of the saliences will be, in DI's view, level or rising tones, a simple computation would be sufficient to give a rough overall total of level and rising tones for the whole text. But, crucially, you would not know where precisely they would occur in the transcription, and no translation software can help in this case. If you were a researcher interested in what happens at particular moments in a recording, you would have to proceed with extreme caution.

More worryingly, the differences between descriptions are mirrored by differences between 'expert' transcribers using the same description. The level of disagreement between independent 'first shot' transcriptions can be as high as 35% (cf Knowles 1991:154; Cauldwell 1994:110). If this low level of reliability is what one can expect from expert transcribers using the same description, it should make us think twice about whether we should spend the time and energy inputting intonation transcriptions at all.

This low figure for reliability can be much improved (up to 90%) by employing a lengthy procedure of independent transcription followed by discussion to produce 'considered transcriptions'. But producing considered transcriptions is time-consuming, and therefore expensive, and is clearly impractical for the large amounts of spoken data that is now going into computer corpora.

If we consider the likely circumstances of transcription, the picture gets bleaker. It is likely that such input would be a single-person unrevised

21

transcription by a non-expert who has had little training in the transcription process, with even more serious consequences for reliability than are indicated by the figures above.

The quality of the recordings discussed above was good: though old, they are of studio quality; additionally the speech events were largely scripted and intentionally tidy, with no overlapping speech. Spontaneous speech is likely to be far less well recorded and far less tidy, and would consequently have many cases of uncertainty. All these factors would lower the level of inter, and intra-description reliability quite considerably.

As mentioned earlier, every transcription an act of interpretation of perceptual phenomena. Because descriptions of the systems of intonation differ in their theoretical standpoints and utilise different perceptual apparatuses the resulting transcriptions are likely to be incompatible in ways that no simple computation can compensate for.

REFERENCES

Brazil, D. 1984. "The Intonation of Sentences Read Aloud." *Intonation, accent and rhythm: Studies in discourse phonology.* ed. by D. Gibbon & H. Richter 46-66. Berlin: Walter de Gruyter.

Brazil, D. 1994. *Pronunciation for Advanced Learners of English.* Cambridge: Cambridge University Press.

Brazil, D. 1995. *A Grammar of Speech.* Oxford: Oxford University Press.

Brazil, D. 1997. *The Communicative Value of Intonation in English.* [2nd Edition]. Cambridge: Cambridge University Press.

Cauldwell, R. 1993. *Evaluating Descriptions of Intonation: A comparison of Discourse Intonation and Systemic Intonation.* Unpublished typescript. EISU: The University of Birmingham.

Cauldwell, R. 1994. *Discourse Intonation and Recordings of Poetry: Philip Larkin reads 'Mr Bleaney'.* Unpublished doctoral dissertation. Birmingham, The University of Birmingham.

Crystal, D. 1969. *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.

Halliday, M. A. K. 1970. *A Course in Spoken English: Intonation.* Oxford: Oxford University Press.

Halliday, M. A. K. 1994. *An Introduction to Functional Grammar.* [2nd

Edition]. London: Edward Arnold.

Knowles, G. 1991. Prosodic labelling: The problem of tone group boundaries. In S. Johansson & A. Stenström (Eds.), *English computer corpora: selected papers and research guide* (pp. 149-163). Berlin: Walter de Gruyter.

Leech, G., G. Myers, & J. Thomas. 1995. *Spoken English on computer: Transcription, mark-up and application*. Harlow: Longman.

O'Connor, J.D. & G.F. Arnold. 1973. *Intonation of colloquial English*. [2nd Edition]. London: Longman.

Roach, P., & S. Arnfield 1995. Linking prosodic transcription to the time dimension. *Spoken English on computer: Transcription, mark-up and application*. ed. by G. Leech, G. Myers, & J. Thomas. 149-160 Harlow: Longman.