QUOTE - UNQUOTE?
The role of prosody in the contextualization of reported speech sequences

Gabriele Klewitz / Elizabeth Couper-Kuhlen

Konstanz
July 1999

In recognition of the enthusiasm he has brought to all aspects of the study of spoken verbal interaction, we dedicate this series to Professor Dr. Aldo di Luzio, University of Konstanz.

Prof. Dr. Elizabeth Couper-Kuhlen
Dr. Susanne Günthner
Universität Konstanz
FG Sprachwissenschaft
PB D 180
D-78457 Konstanz

Prof. Dr. Margret Selting
Universität Potsdam
Institut für Germanistik
Postfach 60 15 53
D-14415 Potsdam

Prof. Dr. Peter Auer
Albert-Ludwigs-Universität Freiburg
Deutsches Seminar I
Postfach
D-79085 Freiburg i. Br.

Abstract

This paper investigates how speakers of English can use the prosodic design of utterances to identify parts of these utterances as instances of reported speech. We will show that prosodic changes can function like quotation marks in written texts by clearly delimiting left and right hand boundaries of the reported sequence. In the majority of cases, however, prosodic changes do not coincide with the boundaries of reported speech but occur nearby, functioning like a 'frame' for the interpretation of a sequence as reported or even only as a 'flag' attracting attention and inviting the listener to actively (re-)construct the corresponding boundaries. Our data analysis also provides evidence for the use of prosodic designs to typify a figure in different roles, which - due to their unique 'prosodic design' - can be presented without any verbalized projection of upcoming reported speech, once they have been introduced. This is due to the 'referent-tracking' nature of some prosodic designs of reported utterances.

Keywords

Reported speech, prosody, contextualization, spoken language, oral quotation, framing, referent-tracking

Konstanz, March 1999

0 Introduction

Discourse - whether spoken or written - is by default understood to be authored by, or to be in the voice of, the person speaking or writing now. When speakers or writers wish to report the speech or thoughts of another person, or when they report words or thoughts of their own entertained at a time other than the moment of speaking, they typically mark them as such: that is, they produce the speech or thought as *reported speech*.[1] The question we shall address here concerns the devices which speakers and writers use in producing discourse as reported discourse.[2]

   One of the primary means of marking discourse as reported is lexico-syntactic: if words or utterances are accompanied by a verbum dicendi or other quotative expression, they will be attributable to a different 'voice'.[3] If a verb of saying has been chosen, speakers and writers reporting discourse have the additional option of deictically anchoring personal, locative and temporal references to the reporting or the reported situation. If deictic expressions are anchored partly or wholly to the reported situation, this alone will cue the discourse as reported, in the absence of a reportative phrase. With or without reportative phrases and deictic transposition, however, discourse can be marked as reported via the convention of quotation marks — in writing. Is there an equivalent to quotation marks in spoken discourse? Or, seen the other way around, what practices in speech do quotation marks in writing (attempt to) capture? In addressing these questions, we shall consider first how quotation marks are used in written discourse and then examine quotation in spoken discourse, comparing and contrasting the two.

1. Quotation in written vs. spoken discourse

The following extracts from a novel by Jane Austen display standard practice for representing reported speech in writing:

---

[1] This is not to deny that on some occasions speakers and writers purposively cloud the issue, leaving it ambiguous as to whose 'voice' the discourse should be assigned.

[2] With Tannen (1989) and Holt (1996) we are not concerned with the ‚authenticity' of reported speech but with its production, i.e. with the (re-)construction in Tannen's words or the 're-creation of a conversation' in Holt's words. Speakers use reported speech to 'demonstrate' a communicative event in its form and content (Clark/Gerrig 1990, Bruenner 1991) and we as analysts must ask which features make a sequence of talk sound 'reported' (see also Mayes 1990). Among these, we assume, are a number of non-verbal cues which index speech as reported (see also Romaine/Lange 1991 and Yule 1995). Our focus, however, will be on prosodic devices specifically.

[3] Degrees of syntactic independence may of course vary.

(1)

1	'In what an amiable light does this place him!' thought Elizabeth.
2	'This fine account of him,' whispered her aunt, as they walked, 'is not
3	quite consistent with his behaviour to our poor friend.'
4	'Perhaps we might be deceived.'
5	'That is not very likely; our authority was too good.'

(Pride and Prejudice: 253)

(2)

1	She wished him to know that she had been assured of his absence before
2	she came to the place, and accordingly began by observing, that his
3	arrival had been very unexpected — 'for your housekeeper,' she added,
4	'informed us that you would certainly not be here till to-morrow; and
5	indeed, before we left Bakewell, we understood that you were not
6	immediately expected in the country.'

(Pride and Prejudice: 259)

There are several things to notice about the way reported speech is handled here. (We assume for the moment the validity of traditional categories such as direct, indirect and free indirect forms of reported speech; see Leech and Short (1981) for a taxonomy.)

(i) Quotation marks surround direct reported speech.
So-called direct reported speech (including reported thought) — with expressive properties (e.g. the exclamation in (1) line 1) and syntactic independence — is set off from the reporting context by quotation marks: see (1) line 1 and (2) lines 3ff. By contrast, so-called indirect reported speech — lacking expressive properties and syntactic independence — is not set off typographically: see (2) lines 1 and 2. Thus it is only direct reported speech which has typographical marks signalling its reported nature.

(ii) The boundaries of direct reported speech are marked exactly.
Quotation marks are placed before the first word and after the last word of the stretch of speech being reported. Words within the quotation marks are by convention understood to be reported; words outside them are by convention understood to belong to the reporting speaker. Direct reported speech is thus delimited or demarcated with precisely placed lefthand and righthand boundaries. Moreover, these boundaries are marked in identical fashion.[4]

---

[4] For the sake of the argument, we momentarily ignore the fact that some fonts use rounded quotation marks facing rightwards and leftwards to mark the beginning and end of reported speech, respectively. We return to this point, however, below.

(iii) The embedded animator [5] of the direct reported speech is as a rule referred to explicitly in a reportative phrase.

For instance, in (1) we find e.g. *thought Elizabeth* (line 1) or *whispered her aunt* (line 2). Such verbal attributions may precede or follow the reported utterance (see (1) line 1), or they may be positioned at specifiable locations within it (see (1) line 2f and (2) line 3). Reportative expressions may, in addition, carry metalinguistic information about how an utterance is presented as having been said (e.g. *whispered* in (1) line 2).

(iv) In extended reported dialogue the identification of the embedded animator may be implicit.

When reporting the speech of two figures in dialogue who have already been introduced, writers sometimes omit verbal indications of who is speaking (see (1) lines 4-5 above). Proper interpretation then depends on the typographical convention of beginning a new line for each new 'turn' in the exchange. The sequence of 'turns' follows the logic of real conversational exchanges.

(v) Different figures in reported dialogue are not individualized typographically.

There is nothing particular about the typographical presentation of, e.g., line 4 in (1) which tailors it to its embedded animator and distinguishes it from the embedded animator of line 5.

In speech — as opposed to writing — it can be assumed that many of the same distinctions need to be made: reported speech must be differentiated from speech itself and different 'voices' must be assignable to different figures (Couper-Kuhlen 1998). Moreover, the same lexico-syntactic and deictic devices for marking reported speech are available. But typographical cues for direct reported speech are obviously missing. Since many typographical conventions — punctuation and italicization, to name two — are commonly believed to stand for prosodic configuration in speech (Halliday 1985: 228), the research question which imposes itself is: *Is the marking of reported speech in conversation done prosodically instead of typographically?*[6] That is, does prosodic marking substitute for typographical marking in speech? In the study which this paper reports on, we have investigated this hypothesis empirically, examining

---

[5]   Goffman (1981) describes three different roles for speakers, with 'animator' being the "individual active in the role of utterance production" (Goffman 1981: 144), 'author' being "someone who has selected the sentiments that are being expressed and the words in which they are encoded" (ibid.), and 'principal' being "someone whose position is established by the words that are spoken [...] whose beliefs have been told [...] (ibid.). Of these three, only the first is necessarily held by the person who mouths the speech. The second and third capture different social roles which can be held by the animator simultaneously or by a separate person. We concentrate here on the 'animator' role, i.e. on the one who is presented as the primary producer of an utterance. The figure (of a story) to whom a sequence of reported speech is attributed, i.e. "a figure in a statement who is present only in a world that is being *told about*" (Goffman 1981: 149, our italics) will be referred to as an 'embedded animator' (ibid.) Multiple embeddings are of course possible. See also Hanks (1990:199ff).

[6]   Cf. also Yule (1995:187).

hundreds of instances of reported speech in conversational English.[7] On the assumption that oral quotation would be cued in one way or another, we first identified passages of reported speech holistically, i.e. using our native/near-native intuitions as competent members of the speech community involved. These passages were then analyzed prosodically, first on an auditory basis and subsequently with acoustic support. In a final stage, correspondences were sought between reported speech and prosodic marking.

Before discussing the results of our investigation, a word is in order on our notion of prosodic marking. Prosodic marking can be identified at different levels of analysis. To take, for instance, a relatively low level: a stressed syllable can be thought of as 'marked' with respect to neighboring unstressed syllables if it is differentiated from them by virtue of higher pitch, greater loudness or greater length. Even a pitch glide can be thought of as 'marked' because it is a dynamic rather than a static tone. But this is not the kind of prosodic marking we have in mind. Conversational reported speech tends to involve whole utterances - not syllables but rather words, phrases, clauses, etc. presented as turn-constructional units.[8] Therefore, for present purposes prosodic marking can be expected to be global rather than local, i.e. to involve departures from norms or expectations holding for pitch configuration, loudness and timing at the level of the intonation phrase or beyond.[9]

---

[7] Our corpus consists primarily of audio recordings of informal conversations between friends and family members and of telephone conversations from radio phone-in programs.
[8] There are of course exceptions to this general tendency, as can be seen in Ex. (17) below.
[9] In our transcription system such global prosodic features are notated in angled brackets before a line of transcript rather than within the line.

2. Prosodic marking of reported speech

Examining the materials at hand, the first observation to make - and one which supports the hypothesis that conversational reported speech is indeed marked prosodically - is that it is often accompanied by a noticeable shift of pitch register or range. To take a rather typical case:[10]

(3) Feeding the neighbor (T92 06:05)[11]
A group of friends talking about 'old times', including once when Eve stayed with Antonella for a couple of weeks.

```
1       Antonella:    have we Ever treated you BADly Eve ,
2       Eve:          [no .
3       Antonella:    [while you were STAYing here ,
4       Eve:          no ,
5                     Even when i moved [next DOOR ,
6       Ruth:                           [(      )
7       Eve:          even when I moved next DOOR -
8                     antonella would show up with the JAR going-
9            f₁: <<h> ↑WE made this yEsterday;>
10                    [and erm ((laughing))
11      All:          [hmhmhm
12      Eve:     f₁: <<h> there ↑ISn't enough for two people
                 f₁: left of this
13               f₁: so ↑YOU can have it>
14                    so we're getting all these like eXOtic
                      ↓like gorgonZOla sauce for my [pAsta .
15      Ruth:                                       [wow
16      Eve:          i nEver would have attEmpted to mAke;
                      (1.0)
17      Ruth:         did you lIke it
18      Eve:          t was nIce
```

In this passage Antonella has just served dinner and Eve has (jokingly) pretended to be astonished at how good it is. Antonella responds with a teasing remark implying that she has always treated Eve well (lines 1 and 3). Eve now agrees with Antonella's assessment and expands on her agreement with an exemplary story, told to Ruth, about how Antonella regularly brought food over to her even once she had moved away (lines 5-13). Antonella's words (f1) when delivering the food are projected by Eve with *going* (line 8). There follows a stretch of talk prosodically set off from the surrounding talk by a shift to high register.[12] The register shift

---

[10] In this and the following transcripts, we adopt the convention of showing our interpretation of where reported speech occurs by placing an indexed 'f' (for 'figure') in a 'voicing column' immediately to the left of the line itself. Absence of 'f' means the words are attributable to the speaker indicated in the 'speaker column'.

[11] The transcription conventions employed here are based on those of GAT (Selting et al 1998).

[12] According to Cruttenden (1986) one speaks of a shift in register, when top- and baseline of the pitch range used by a speaker are shifted so that the whole stretch of utterance is perceived as higher or lower than surrounding talk.

begins on *we* and ceases after *yesterday* in line 9. And it begins again on *there* in line 12 and ceases after *it* in line 13. These two stretches of speech, delimited by prosodic marking, are heard as Antonella's words, whereas line 10, lacking high register, is heard as being in Eve's own voice. In this fragment then, stretches of reported speech which would be enclosed in quotation marks in writing are marked off by shifts into and out of high register.[13] The register shifts begin and end precisely where the quotation marks would be placed in writing.[14] See the following graph[15]:
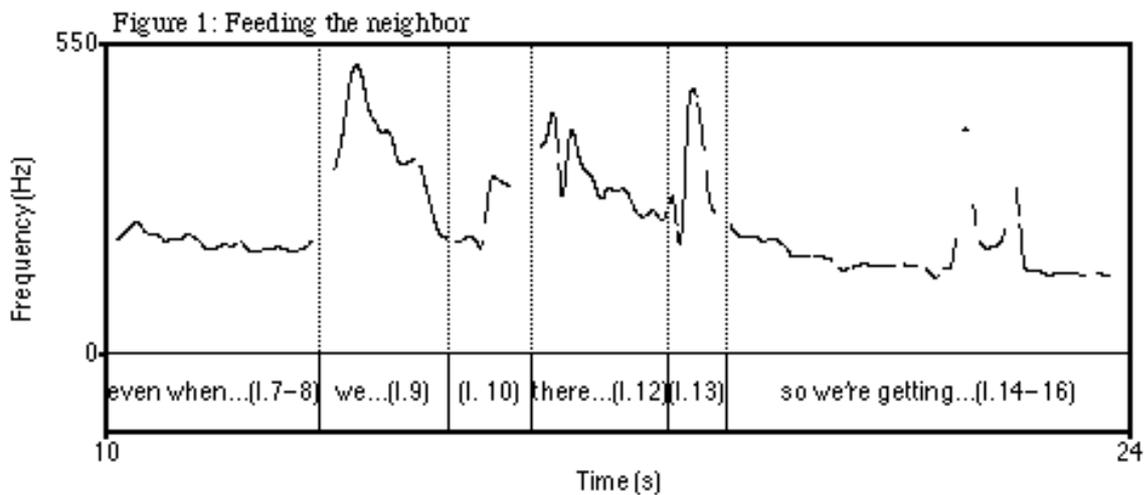
Not only register shifts but also changes in volume are found coinciding with stretches of conversational reported speech. For example:

(4) *The Twit (T93 05:40)*

```
        1       Eve:            they knew I couldn't STAND him;
        2                       .hh <l(h)ike,
        3                       you know I'd come BACK after lunch
                                going>              <laughing>
        4               f₁:  mmm
        5                       <<h, n>they'd be like>
->      6               f₂:  .hh <<f> ´HOW i:s - >
        7                       and=I'd=say
```

---

[13] Our corpus also contains cases of register shift to low coinciding with reported speech. For an example see (6) below.

[14] Note the similarity between rounded quotation marks facing in opposing directions and switches into and - at the end of the reported sequence – out of a chosen set of prosodic parameters.

[15] The graphs in this paper have been created with the help of Praat 3.8, developed by Paul Boersma (for more information see http://www.fon.hum.uva.nl/praat/). Whenever a pitch contour is shown it has been smoothed to make the visual presentation more closely match our auditive perception.

```
8                           f₁:  nga nga nga
9        Antonella:      'hm'hm
```

In this excerpt Eve is telling her friends Ruth and Antonella about an annoying co-worker at her summer job. In lines 3-8 she re-enacts a typical dialogue between herself (f_1) and a chorus of women colleagues (f_2), who are teasing her about him. The words and expressions of Eve and the chorus are projected verbally with *I'd come back ... going* (line 3), *I'd say* (line 7) and *they'd be like* (line 5) respectively. The voice of Eve the figure is animated primarily with non-verbal expressive sounds such as *mmm* (line 4) and *gna gna gna* (line 8), whereas the chorus' voice is animated with a (truncated) question *how is* - (line 6). It is the latter stretch of reported speech which coincides with a sudden shift to loud volume. Eve resumes her normal volume again when she continues the narration in line 7.

Figure 2: The Twit
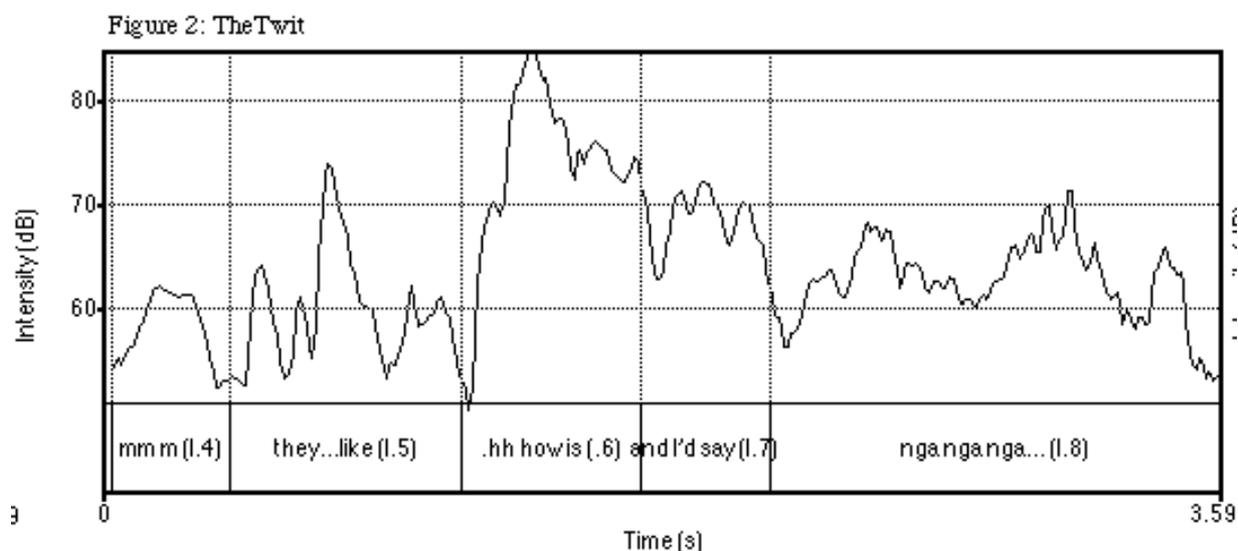
In addition to passages of reported speech which coincide with a simple increase in overall pitch or volume, we have also found cases in which there is a shift to perceptually isochronous timing (see Auer/Couper-Kuhlen/Mueller, in print) on a stretch of direct reported speech:

(5) *Hospital inquiries (T12a 47.53)*

```
1      Joan:           they tOOk him to the HOSpital and uh -
2                      CAroline said -
3                      on the wAy through the bUshes ;
4                      the bUsh caught him ;
5                      <<p,l>and went like >
6                      you know
7      Anne:           hm
8      Joan:           yeah
```

10

```
        9                         he had a - (.)
        10                        lIne across here ,
        11                        she said
->      12              f₁: /BOY did they ever /
                        f₁: /Ask us            /
                        f₁: /quEstions
        13      Jan:          ha ha ha ha
```

In this excerpt Joan is telling friends about a sledding accident in which her grandson *went into the bushes* and *cut his eye.* Her report is based on that of Caroline, the boy's mother, who accompanied the child to the hospital. Joan projects Caroline's words (f_1) describing her experience at the hospital with *Caroline said* and *she said* (line 2 and line 11). The talk that follows line 2 is heard as an insert providing background information necessary to appreciate the punch line, *boy did they* [i.e. the hospital crew] *ever ask us questions.* The latter is presented subsequent to line 11, in a stretch of speech which - in contrast to surrounding talk - is remarkably regular due to the fact that the accented syllables on *boy, ask* and *questions* are placed at equal intervals in time. The rhythmic pattern begins with the first word of Caroline's report *boy* and ends with its last word *questions* (line 12). Reported speech here is thus coextensive with a shift to perceptually isochronous timing of the accented syllables.[16]

Figure 3: Hospital inquiries
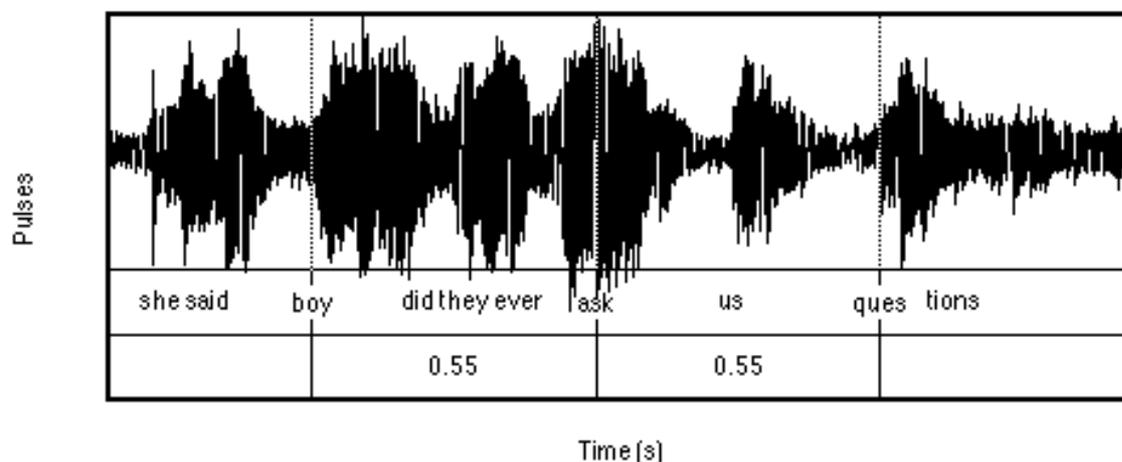
When the speech of more than one figure is being reported, separate prosodic marking may coincide with the different voices. The following example demonstrates a case in which separate

---

[16] The regular intervals marked off in this and the following wave forms extend from vowel onset to vowel onset of the accented syllables in questions; the absolute duration of each interval is indicated below it in hundredths of a second. See Couper-Kuhlen (1993) for a discussion of why this kind of measurement is used and of how much durational variability can be tolerated within the bounds of perceptual isochrony.

perceptually isochronous patterns are used for the speech of different re-enacted figures (Bill the figure in line 2-3 and his ex-wife Jill in line 5):

(6) *Divorced Parents I (T12a.2, 1:44:44)*

```
       1      Bill:          w- what what i told A:NN is,
->     2             f₁: <<h>right    /↑nOw ; (.)               /
       3             f₁:          /jIll:=is sEttin' me    /
                     f₁:          /Up ,>
       4             she's like
                     f₂: <<l> w-d-
->     5             f₂:      /bIll is    /
                     f₂:      /nOt        /
                     f₂:      /shOwin a   /
                     f₂:      /lOt of     /
                     f₂:      /INterest,>
```

In this excerpt Bill is telling his sister about a discussion he had with his ex-sister-in-law Ann. He is complaining that his ex-wife Jill is 'setting him up', i.e. creating the (false) impression that he is not paying enough attention to their young daughter. When Bill re-enacts the conversation he had with Ann, he projects his own words with *what I told Ann is* (line 1); the reported speech which follows (f₁, lines 2-3) is very rhythmic, with perceptually isochronous accents on *now, Jill,* and *up.*[17] When Bill resumes narration and now introduces the figure of Jill (line 4), rhythmic delivery is discontinued. Jill's speech (f₂), however, is re-enacted with perceptually isochronous timing of the accents on *Bill, not, showin', lot* and *interest,* in a new and different rhythmic pattern, characterized by short intervals and a gradual increase in tempo, as illustrated in the following graph:

---

[17] Despite the absolute difference in acoustic duration between the two intervals involved (see Figure 4), the proportional difference is only 20%, which is well within the bounds of perceptual isochrony in speech (Couper-Kuhlen 1993).

Figure 4: Divorced Parents
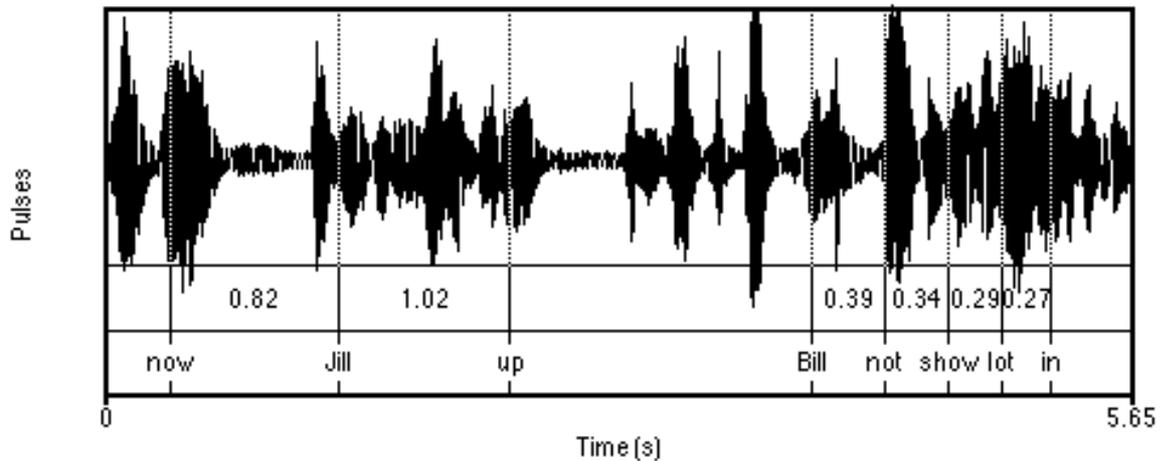
The above excerpt also illustrates that passages of reported speech may be marked not just by a shift in one prosodic parameter (e.g. pitch as in (3), loudness as in (4) or rhythm as in (5)) but by multiple prosodic shifts occurring simultaneously in a cluster. In (6), the first bit of voicing (Bill the figure, f_1) not only has a prominent isochronous pattern but also a shift to high register. The second voicing (Jill, f_2) has its own isochronous pattern and a shift to low register. Clusters of prosodic shifts as marks of reported speech are actually more common than simple ones in our corpus, and they are often accompanied by changes in voice quality. The fact that multiple prosodic and paralinguistic shifts co-occur is indicative of the nature of the contextualization process involved (Gumperz 1982, Auer/di Luzio 1992). The shifts contexualize the speech as reported, that is they serve as quasi-iconic signs of a shift in indexical ground (see also Hanks 1990:205ff).

Yet the clusters of prosodic cues which accompany reported speech are not random collections. They are first of all not local but global ones: register, volume and rhythm shifts — as well as changes in pitch range and speech rate[18] — which are relevant for the domain of the intonation phrase and/or a sequence of intonation phrases. Moreover, judging from the frequency of certain combinations, there appears to be a hierarchy of preference among the prosodic cues. For instance, pitch shifts (with or without a change in voice quality) are almost always part of a combination format. Shifts to lento and allegro rarely occur alone as markers of reported speech. Yet as we point out below, which specific prosodic format is used ultimately depends on the circumstances of the context in which the reported speech is situated.

---

[18]   A speech rate shift coinciding with reported speech will be observed in Ex. (7), a pitch range shift in Ex. (11) below.

Another point needs to be made about the prosodic delimitation of quotation in spoken discourse. Although all preceding examples have contained verbally projected reported speech, it goes without saying that prosodic marking also coincides with non-projected quotation. For instance:

(7) Divorced parents II (T12a.2, 1:45:20)

```
        1       Bill:           what she's DOin' is -
        2                       she's settin' me up -
        3                       she's also settin juliETTE up .
->      4               f₁: <<l, len>   /lOOk how much   /
                        f₁:             /fUn you         /
                        f₁:             /hAve with your  /
                        f₁:             /AUNT.> (---)
        5                       and that's why I told a- Ann ,
        6                       i said i rEAlly think -
        7                       she's sEtting me ´UP ;
```

In this excerpt Bill is complaining that his ex-wife Jill is not only 'setting up' him (e.g. by presenting him as an uncaring father, see (6) above) but is also 'setting up' their young daughter Juliette (line 3). Without further introduction Bill now animates[19] Jill's voice (f_1), reporting words she might use in conversation with the child to 'set her up': *look how much fun you have with your aunt* (line 4).[20] The animation of Jill's voice coincides with three prosodic shifts: low register, isochronous rhythm and slow tempo. As of line 5 there is a shift back to the prosodic dimensions which characterized Bill's talk prior to line 4. These three prosodic shifts 'tag' the passage in question as attributable not to Bill but to someone else.[21]

If there is no introductory verb of saying, prosodic marking may be the only external sign of the quotative nature of a stretch of speech. Yet the results of our survey do not suggest that there is more prosodic marking with non-projected than with verbally projected reported speech. Nor do they indicate that multiple shifts are more frequent in one case than in the other. The choice of prosodic formatting for conversational reported speech appears to depend less on type of verbal projection than on such factors as how the figures are being characterized and what stance the speaker is taking to the speech and the figures in question (see below).

---

[19] Bill the narrator attributes reported speech to his figure Jill, thus producing what Goffman called 'embedded animation'. For the sake of formulation we will not stress the embeddedness of this kind of animation for each extract and continue instead with the simpler term 'animation'.

[20] Since Juliette's aunt is being considered as a potential foster parent, the implication is that Juliette has more fun with her aunt than with Bill her father and should therefore choose to stay with her.

[21] Who the words belong to must be inferred from their content, in conjunction with the way they are configured prosodically (see below).

The fact that prosodic marking coincides with passages of reported speech in conversation is not happenstance. Rather, the prosodic shifts can be thought of as accomplishing something, namely marking — and in the cases examined so far — delimiting one's own territory or speech from someone else's. In this delimitative sense the prosodic shifts function like quotation marks in written discourse: where a prosodic shift begins, quoted speech begins; where this shifts ends, the quoted speech ends. But with this kind of prosodic marking, the material under quotation is continuously marked (which makes it particularly apt for spoken language), whereas with typographical marking only beginnings and ends are signalled.

The relation between prosodic marking and conversational reported speech — as sketched so far — appears rather straightforward. But in fact the situation is much more complex. For one, prosodic marking is not restricted to <u>direct</u> reported speech in conversation. Consider, for instance, the following excerpt:

(8) Musty basement (Tape 12a:187)

```
           1      Ann:          Christian was r-
           2                    in the- in the newspaper yesterday
           3                    and they had-
           4                    saying that
->         5                    f₁: /hOUses are  /
                                f₁: /nOw being   /
                                f₁: /^  (.)      /
                                f₁: /wAshed a-   /
                                f₁: /wAy.
           6      Deborah:      yeah
           7      Ann:          before they were s-
           8                    THERE
           9                    but full of water
           10                   and now they're (.) being
->         11                   f₁:? wAshed awAy,
```

Ann and her mother Deborah are talking about the severe flooding which has just recently taken place in the Des Moines area. In lines 1ff Ann reports what the newspaper which her husband Christian has read said (f_1). In doing so she uses a form of reported speech canonically known as 'indirect speech', signalled here by the presence of syntactic subordination to a verbum dicendi (*saying*) and the use of *that* as a complementizer (line 4). Yet the report itself is given strong expressive marking[22] with the perceptually isochronous timing of the accented syllables in *houses, now, washed, away*.[23]

---

[22] There is also a paralinguistic change to an 'awe-stricken' voice.
[23] Note the presence of a 'silent beat', represented with /^/, intervening between *now* and *washed* (for further discussion of this phenomenon see Couper-Kuhlen (1993) and Auer/Couper-Kuhlen/Mueller (in print)).

Figure 5: Musty Basement
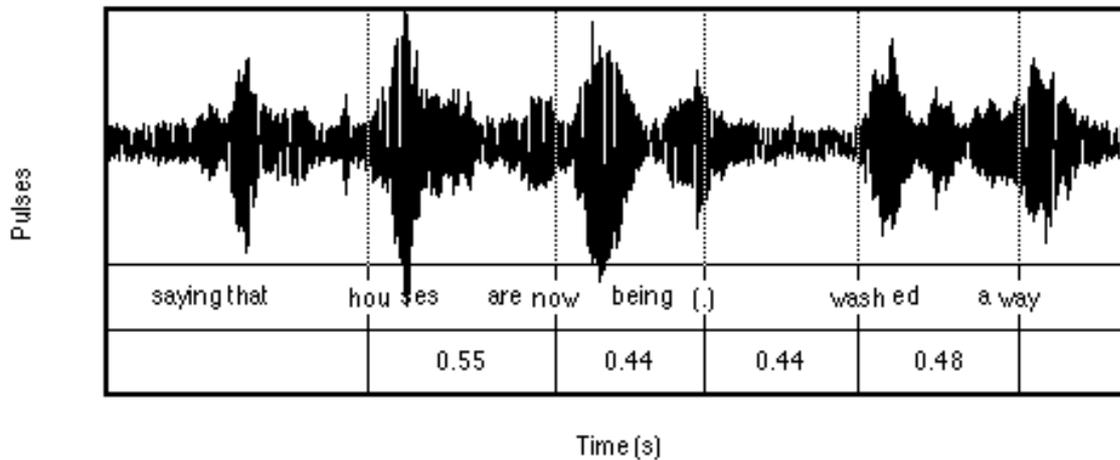
Prosodic marking of 'indirect' speech in conversation is by no means a rare phenomenon. In fact, it is actually rather widespread, particularly in high-involvement talk. This points to an important difference between prosodic marking in speech and typographical marking in writing: whereas quotation marks are reserved for 'direct' speech only, prosodic marking can be found with all forms of oral reporting. As Günthner (1997a, 1997b, 1997c) has shown, expressive prosodic marking cuts across the canonical, grammatically based distinctions between 'direct' and 'indirect' speech, suggesting that these categories as traditionally defined are not necessarily the relevant ones for spoken interaction.

But there are more ways than just this one in which prosodic marking in speech does not correspond to quotation marks in writing. In addition, not all (direct) reported speech in conversation is prosodically marked. This is true both in a somewhat trivial sense and in a more important way. To illustrate the more trivial sense first:

(9) *Beverly (T39 24:00)*

```
        1       Jane:           Beverly said
                        f₁: she had a ↑WOnderful TIme ,
                            [.hh
        2       Sue:            [are y- SURE ,
        3       Jane:           <all> i was like ;>
->      4               f₂: do you reMember STANding up and sayin=eh (.)
        5                   <all, p, l> i didn't Ask her > ,
        6       Sue:            haha
```

16

In this fragment Jane and Sue are talking about a dinner party they had the night before to which Beverly was invited. Jane has just run into Beverly prior to this conversation and now reports what Beverly said, namely (f$_1$) *she had a wonderful time* (line 1). This stretch of (indirect) reported speech has no special prosodic marking. Line 4, however, is a direct quotation of Jane's response to Beverly, introduced with *I was like* (line 3). Jane's reported words (f$_2$) *do you remember standing up and saying* have no special prosodic formatting either: they are said in her normal voice. Significantly, however, this 'normal voice' is not the one used to introduce the quote; *I was like* has fast (allegro) speech rate. The reported speech of line 4 thus contrasts prosodically with the reporting phrase because it is non-allegro. Moreover, it also contrasts prosodically with what follows, namely the return to the narrator's voice in line 5:[24] this line is prosodically marked with allegro speech rate, piano volume and low pitch register. Therefore, what we have here is a reversal of markedness: although the passage of reported speech is not prosodically marked in the literal sense, it is 'marked' in the reverse sense of being unmarked in a marked context.

But more importantly, even when direct reported speech is embedded in a prosodically unmarked context, it may remain unmarked:

(10) *New Job (Y-Coll 13:18)*

```
->      1    Eve:          and I was there (kinda) going
                     f₁: you REAlize I haven't done this
                     f₁: and I haven't a CLUE ;
        2            f₁: you know ,
        3                and he went -
        4            f₂: <<h> Oh well maybe you should pay >
                     f₂: Undergraduate fees as WELL as pOstgraduate
        5    Sam:         [heh heh heh
        6    Eve:    f₂: [<<h> cause YOU'LL be learning tOO ;>
        7                <<all>I was there kind of going
                     f₁: <<l> nO ,>>
        8    Antonella:    ((laughing))
```

In this excerpt Eve is telling her friends about a job she once had as a teaching assistant. She re-enacts one of the first meetings she had with the head of the department, at which it became clear that she would have to teach a subject she had not studied herself. Eve's own voice (f$_1$) is animated in lines 1ff and 7, that of the head of department (f$_2$) in lines 4 and 6. Yet although she introduces her own words in line 1 with *going* and reports them 'verbatim', they have no

---

[24] In context, line 5 is heard as a response to Sue's question *are y- sure* (line 2); see below.

particular prosodic marking at all.[25] It is not until the department head's voice is animated that any kind of prosodic marking is used.

Figure 6: New Job
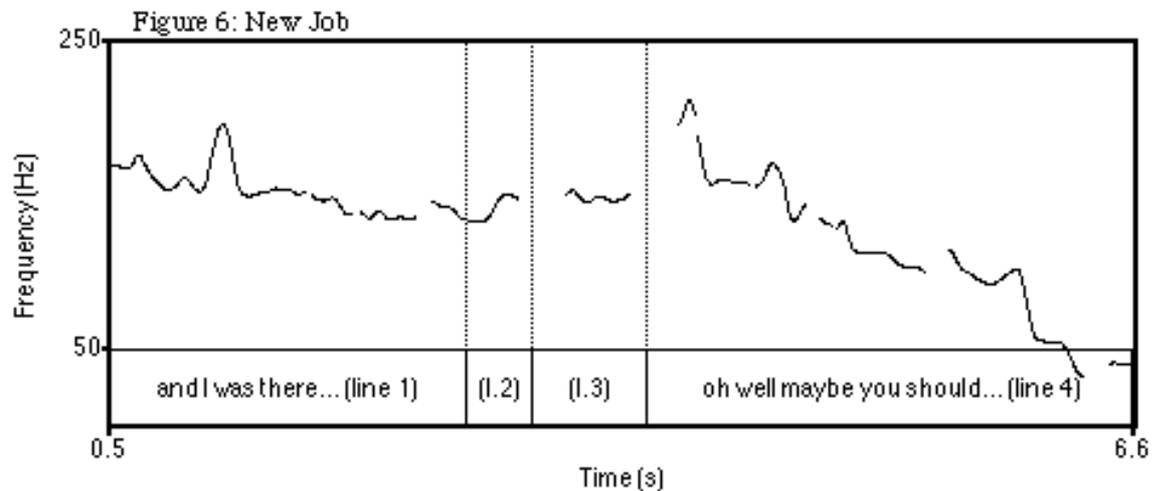
This fragment demonstrates that speakers have the option of marking passages of (direct) reported speech prosodically or not. In the case at hand it could be to the narrator's advantage to present the protagonist (herself) as behaving in a normal 'unmarked' way and to contrast the behavior of her department head, the antagonist, as 'abnormal' and 'marked'.[26] In other words, no prosodic marking may be a stylistic choice for speakers who are reporting speech events.[27]

Conversely, it should be pointed out, prosodic marking may be a stylistic choice even when no speech events are being reported at all. To see this, reconsider example (9).

(9) *Beverly (T39 24:00)*

```
        1       Jane:           Beverly said
                        f₁: she had a ↑WOnderful TIme ,
                            [.hh
        2       Sue:            [are y- SURE ,
        3       Jane:           <all> i was like ;>
        4                   f₂: do you reMember STANding up and sayin=eh (.)
->      5                       <all, p, l> i didn't Ask her > ,
        6       Sue:            haha
```

Line 5 *I didn't ask her* has a cluster of prosodically marked cues, allegro speech rate, piano volume and low pitch register. Yet it is not interpretable as the reported speech of any of the figures in Jane's story. Instead this line belongs to Jane the narrator; she is heard as responding to Sue's prior (overlapped) query *are y- sure* (line 2). In this case the prosodic marking contributes to the interpretation that her remark is an aside, unrelated to the main business of recounting Beverly's reaction to the dinner party: it does not mark an oral quotation.[28]

In sum, although stretches of conversational reported speech are frequently marked prosodically, it would be an overstatement to claim that prosodic marking is used systematically as a sign of reported speech in talk the way quotation marks are in texts. For one, not only direct but also 'indirect' reported speech is often prosodically marked. Moreover, oral quotations may be cued as such by virtue of being prosodically unmarked if they are embedded in a prosodically marked environment. And more importantly, quotations in speech may not receive any kind of special prosodic formatting at all. The lack of marking may be a conscious stylistic choice on the part of the speaker. Unlike in writing where one kind of quotation mark (single, double, straight or round) is chosen for all instances of reported speech, the prosodic marking of several instances of reported speech by one speaker may vary from one reported character to another and some instances may be unmarked altogether. Finally, in contrast to quotation marks, prosodic marking is not restricted to quotation but also occurs elsewhere, signaling other structural and expressive dimensions of talk.

3. Prosodic framing vs. flagging of reported speech

If prosodic marking does not correspond systematically to written quotation marks, can patterns nevertheless be established for the way in which it signals conversational reported speech in those instances where it is used? The prosodic formatting of reported speech in fragment (10) above is worth looking at again in this respect.

(10) *New Job (Y-Coll 13:18)*

```
        1       Eve:          and I was there (kinda) going
                        f₁: you REAlize I haven't done this
                        f₁: and I haven't a CLUE ;
        2               f₁: you know ,
        3                   and he went -
->      4               f₂: <<h> Oh well> maybe you should pay
                        f₂: Undergraduate fees as WELL as pOstgraduate
        5       Sam:          [heh heh heh
```

---

[28] Similarly the prosodic marking of line 5 in example (5) does not mark an oral quotation.

```
6       Eve:        f₂: [<<h> cause YOU'LL be learning tOO ;>
7                        <<all>I was there kind of going
                    f₁: <<l> nO ,>>
8       Antonella:   ((laughing))
```

Notice that when Eve animates the voice of the head of department (f₂) she suddenly shifts to high register on *oh well* (line 4). But this high register is not maintained throughout the rest of the speech attributed to the department head. Instead, Eve's pitch gradually drops over the course of the next phrase *maybe you should pay undergraduate fees as well as postgraduate,* so that by the time she has reached *postgraduate* her register is back to normal.

Figure 7: New Job (line 4)
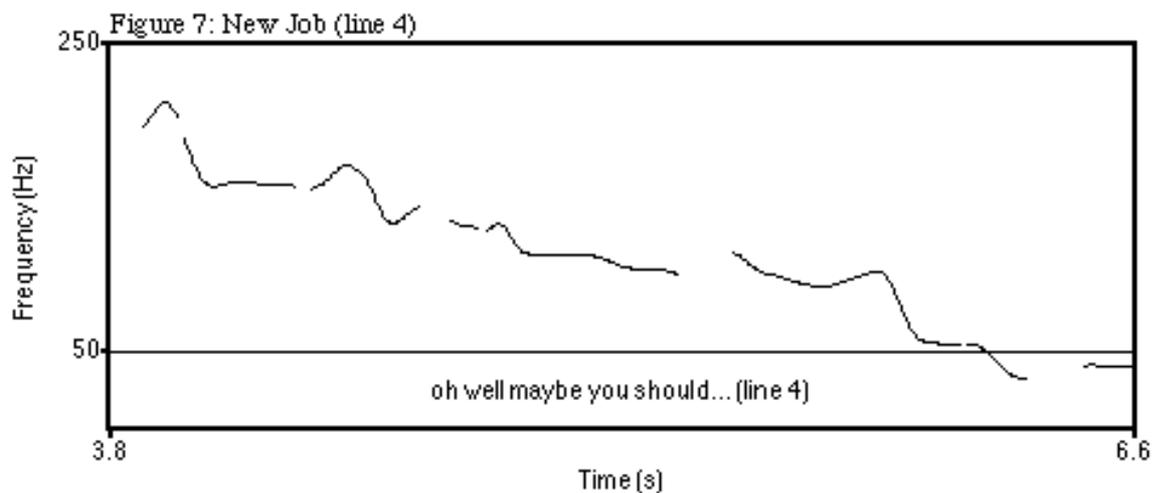
With the next intonation phrase, also attributable to the department head, *cause you'll be learning too* (line 6), she switches back to high register again. Thus fragment (10) suggests that it may not be necessary for prosodic marking to be sustained throughout a stretch of reported speech. Here reported speech is merely framed by prosodic shifts at the beginning (shift into high) and the end (shift away from high). With prosodic framing there is more similarity between typographical marks and prosodic 'marks' for reported speech than with sustained prosodic marking.

Yet in contrast to typographical marking, the prosodic boundary marks of reported speech are not always identical. This is because the prosodic formatting of a voice may 'evolve' during the stretch of speech being reported. When this happens, the left- and righthand boundaries may end up being different, as in the following example:

20

(11) Homophobia(T33, Dat 1:00:45)

Talk show on homosexuality. The guest, a former Chief Rabbi, has just remarked that the existence of a homosexual gene could be used to single out and treat homosexuality. This remark has been sharply criticized by the other guests.

```
1      Rabbi          <<all>that there was a> ge/NEtic /FActor
                      re/SPONsible for er er uhm homosExual
                      predisposItion ;
2                     that was a statement that I didn't make
3                     I was not a party to it
4                     it /hAppened to /↑HIT the /nEws .  .hh
5                     and was used by homoSExuals ,
6                     and homosExual GROUPS ,
7                     to say ,
8          f₁:        <<w>you SEE ?
9          f₁:        it's just a <<all>nAtural> conDItion ,
10         f₁:        we're not re↓SPONnsible for it> ,
11         f₁:        <<low>it's er something that nature er has CAU:SED
                      in Us> .
12                    <<h>I do not DOUBT> that ;
13                    I don't believe by the way that that discovery er
                      is
14                    <<all>anywhere> near correct
```

The rabbi, in clarifying his statement, attributes a longer sequence (lines 8-11) of reported speech to *homosexuals and homosexual groups* in general (lines 5-6). Although he starts the reported sequence with extra wide pitch range, he switches to normal range and low register in line 11:

Figure 8: Homophobia
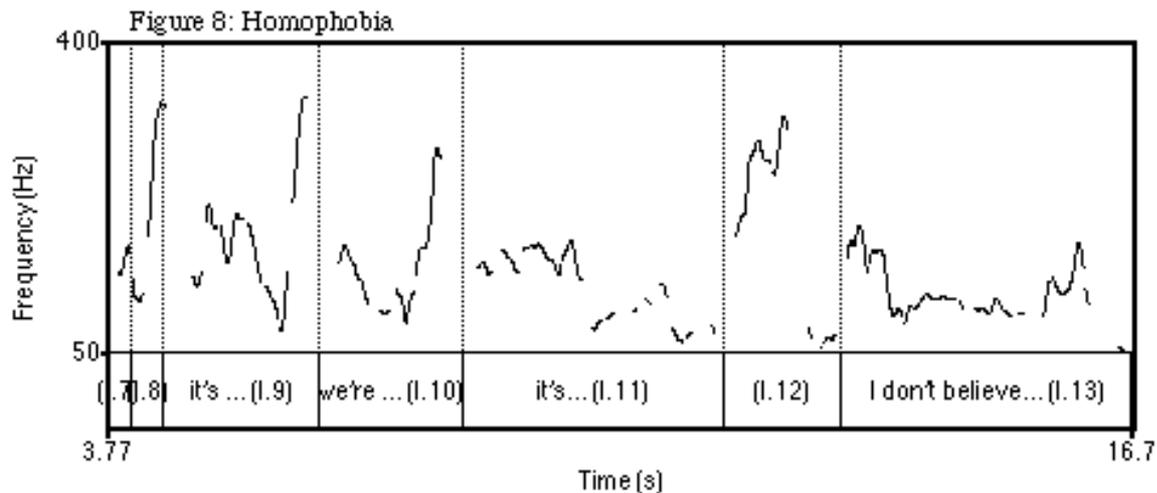
Prosodic changes within a passage of reported speech may cue something about the internal structure of the turn. In this case the change serves to set off a contrast in the reasoning attributed to homosexuals by the rabbi. Line 10 is a denial of responsibility for their homosexuality, line 11, a much stronger contrasting statement that nature has caused it.

21

Stretches of reported speech are thus not necessarily monolithic, but may have internal structure which prosodic resources can be used to cue. When this happens the frame which a prosodic shift creates for a quotation may look slightly different at the end as compared to the beginning.

In addition, there are prosodic devices which serve to cue only the left-hand boundary of reported speech: e.g. the rhythmic pause:

(12) Cuz 5

```
        1       Alina:          the mAle athletes were inCREdible,
        2                       but One of the guys,
        3                       his ENtrance,
        4                       on- --
        5                       (.) (h)on to the stA:ge (h) is he,
        6                       h comes sliding across on his ↑HEA:D.
        7                       (.) h and hEctor,
        8                       of course,
        9                       looks over at me and goes,
        10                      heh
->      11              f₁: (.) <<low>I couldn't do that Lina,
        12              f₁: (.) h Otherwise I'd have NO:: hair lEft> .
        13                      heh heh heh
```

Alina's report of Hector's words (f₁) on seeing the male athlete is projected with *Hector...goes* (lines 7 and 9). His actual words, however, do not follow immediately but only after a noticeable pause (line 11). On one level this pause interrupts the articulatory flow of speech and creates a split second of suspenseful delay: what did Hector say? But on another it helps preserve the isochronous rhythm of speech and thus contributes to cohesion between reporting and reported speech:

(12´) Partial rhythmic notation of (12)

```
        7/8                     (.) h and  /hEctor, of course,             /
        9/10                               /looks over at me and goes, heh /
->      11                                 /^ (.)                          /
                        f₁:                /<<low>I couldn't do that        /
                        f₁:                /Lina, (.) h                     /
        12              f₁:                /Otherwise I'd have              /
                        f₁:                /NO:: hair lEft> .
```

The pause before Hector's *I couldn't do that Lina* (line 1) is a rhythmic pause (Couper-Kuhlen 1993), creating a prosodic boundary between reporting and reported speech. Rhythmic pauses are not uncommon before quoted 'lines' in conversational stories, in particular if the latter are carried off as polished performances.

In fact, there is some evidence in the data we have examined that a <u>non-rhythmic</u> pause which precedes a stretch of reported speech introduced as such is marked and triggers additional inferencing. For instance, in the following fragment, the pause which follows the reportative phrase *Dad says* constitutes an arhythmic break with respect to prior talk:

(13) Underwear (Y-Coll 26.40)
Neighbors chatting about Michael Jordan

```
        1      Dora:           i like the commercial he made with his dad
        2                      where they're
                               /lOOking at this    /
        3                      /Underwear, and the/
        4                      /dAd says
->      5                       (.)
        6                      he's   /gOt all these          /
        7                                 /dIfferent colored      /
        8                                 /Underwears, and he     /
        9                                 /sAys                    /
        10          f₁:              /^ does it             /
        11          f₁:              /wOrk,
        12            and he-
        13            and his wife says
        14      f₂: yes it /dOEs!
                                    <<all> and ye- and they /
        15                      /knOw > and the              /
        16                      /fAther says                 /
        17          f₁:              /^ maybe                /
        18          f₁:              /I should               /
        19          f₁:              /trY it!
```

Dora is recounting a Michael Jordan commercial in which father and son are looking at underwear. Once the dad's line has been introduced (lines 3-4), there is a slight pause (line 5). This pause could of course be a rhythmic pause, i.e. the accented syllables of subsequent speech could maintain the beat across it. But Dora pauses too long. In fact, when she does continue with *he's got all these different coloured underwears* (lines 6-8), these words come across not as a quotation of the father but rather as a 'repair': an insert of background information deemed necessary for proper understanding of what is to come. The arhythmic pause thus cues lines 6-8 as being <u>not</u> the reported speech. But it can only do this against the background expectation that reported speech is regularly introduced rhythmically. Notice that the subsequent lines — by the father, the wife and the father again — are all rhythmic in one way or another. The father's two turns (f_1) are both preceded by a rhythmic pause which serves as a left-handed frame for the upcoming quoted speech; the wife's speech (f_2) is used to establish a new isochronous pattern, which sets up a rhythmic frame for the final punch line.

Examples (12) and (13) are revealing in yet another way: the prosodic formatting of the reported speech passages, namely perceptually isochronous timing of accented syllables,

extends beyond the reported speech itself. In both cases there is thus evidence that the prosodic boundaries or frames for reported speech are permeable. This point becomes even more clear from the following excerpt:[29]

(14) Cuz 4

```
      1      Alina:        Jan talked the whole time,
->    2                    (f₁) <<ff,hh,all> in a voice like this.
      3                    f₁: hi (Mar Alina)
      4                    f₁: I'm so happy to see you.>
      5      Lenore:       [ha ha ha
      6      Alina:        [and we're going,
      7      Lenore:       [ha ha ha
      8      Alina:     f₂: [<<len> Go::d,
      9                 f₂: turn the vo::lume down(h),
     10                 f₂: <let me ou:tta here.>>          <nasal>
     11      Lenore:       ha ha ha
```

In re-enacting what happened during her visit to Jan, Alina animates Jan's voice (f₁) using extra high, fortissimo and allegro prosody (lines 3-4). But these prosodic shifts set in with the reportative phrase *in a voice like this* (line 2), thus foreshadowing the upcoming animation of f₁.

   Although foreshadowing the prosodic design of an upcoming 'voice' (termed 'pre-voicing' in Ruehmann 1999) during the introductory reportative clause is the most frequent type of anticipatory pattern, there are others documented in our corpus. For instance:

(15) Cuz 6

```
      1      Alina:        th- the friend that was there with them,
      2                    is this older guy with this young chick.
      3      Lenore:       h
      4      Alina:     (f₁) <<h> and she was like a real pill
                                                  <nasal, breathy>
      5                 (f₁) you know,>
      6      Lenore:       ha ha ha
      7      Alina:     (f₁) <<h> she's sitting there,   <nasal, breathy>
      8                 (f₁) with this hair pulled back,
      9                 (f₁) in a /little /pony /tail.
     10                 (f₁) and she's like sitting there and,>
     11                    he said,
     12                 f₂: <<f, all> I would have been here,
     13                 f₂: but <she was so late.
     14                 f₂: and getting her any place on time,>> <nasal>
     15                    she's going,
     16                 f₁: <<h> well,
     17                 f₁: I had to get ready.>   <nasal, breathy>
     18                    I don't know why.
     19      Lenore:       ha ha ha ha
```

---

[29] We use an f in parentheses in the voicing column below to indicate that the speaker is talking about a figure while at the same time using prosodic or lexico-syntactic features attributable to that figure's 'voice'.

```
20     Alina:          ((snort))
21                     nothing was gonna help her.
22                     no make-up,
23                     no nothing.
24             (f₁) cause she's the little <<h> girl,
                                            <nasal, breathy>
25             (f₁) and he's the older man,
26             (f₁) and he's taking care of me.>
```

The story that Alina is telling has two main figures, the *older guy* (f_2) and the *young chick* (f_1). The man's voice (lines 12-14) is introduced with a reporting phrase *he said* (line 11) and is accompanied by prosodic shifts to forte and allegro. The young girl's speech (f_1) in lines 16-17 coincides with a marked prosodic shift to high register, accompanied by a paralinguistic shift to nasal, breathy voice. Yet these features are not confined to the girl's speech only; they appear whenever she is being referred to. The 'voice' of f_1 with its characteristic prosodic and paralinguistic design is foreshadowed in lines 4-5 and 7-10. Moreover, it is echoed in lines 24-26, once again with high, nasal and breathy voice. The injection of expressive prosody and paralinguistics presented as characteristic for this figure into Alina's report is an example of the narrative device called 'free indirect speech', often believed to occur only in written discourse. Yet this example and many others like it demonstrate that free indirect speech is as much a phenomenon of spoken as of written discourse, provided prosody is taken into account.

In example (15) distinctive prosodic formatting not only anticipates but also echoes the 'voice' in question. Pre- and post-voicing thus create a reciprocal permeation of the left- and righthand boundaries of reported speech. The anticipation or echoing of a prosodic design should not be thought of as a lack of articulatory control on the speaker's part but rather as a strategic device for prefiguring or 'postfiguring' the upcoming reported speech and the way it will be presented. This technique belongs to the general repertoire of devices which speakers have for overcoming the temporality of spoken discourse. By foreshadowing and, on occasion, echoing the animation of a figure's voice, prosodic cues can be used to prime the context for a particular figure and the way it is stylized, thus in effect prolonging its presence 'on stage' and heightening its rhetorical effect.

The phenomenon of prosodic pre- and post-voicing makes it difficult to speak of prosodic framing for reported speech in a strict sense. When it happens, the reported speech which is supposedly being 'framed' is not uniquely 'within' the frame; instead the 'frame', if it is one, is larger and includes non-reported material. For this reason it might be preferable to call this kind of prosodic signalling of reported speech 'flagging', the prosodic shifts being comparable to flags set up in the vicinity of the phenomenon to be signalled. The flags, however, need not mark the whole of the territory in question, nor even its precise boundaries. Instead they seem

to say 'Watch out for something in my vicinity'. In this sense they are not iconic but indexical of a deictic shift in reported speech. Although prosodic flags can be compared to quotation marks in written discourse, they function much more loosely and approximatively, as the above discussion has shown.

4. Referent tracking and prosody in reported speech

There is another sense in which the prosodic handling of reported speech in conversation differs from its typographical marking in texts. Recall that in written discourse, when a dialogue is being reported, there is no typographical distinction made between the speech of different characters, aside from the fact that a new line is begun with each new speaker. In other words it may be difficult to keep track of multiple referents (which figure is speaking now?) in the absence of some verbal tag such as *he said* or *she said*.[30] In spoken discourse, on the other hand, the use of prosodic and paralinguistic marking, framing, or flagging opens up a vast array of resources for individualizing figures. Once a figure's voice has been individualized prosodically, the continued use of this prosodic design will permit referent tracking without verbal tagging, even when the 'cast' of figures is larger than two.[31]

To see how this works, consider the following, rather typical case:

(16) Jane Kerr (Y-Coll 27.05)

```
      1     Jane:        and she said ; (-)
      2                f₁: <<h,all> do you wOrk-
      3                f₁: do you wOrk in apple VALley?>
                                             <breathless>
      4                     and he said
                         f₂: <<l> ´`YEAH,>
      5     All:         h, heh hmm
->    6     Jane:        f₁: <<h, all> do you know jane KERR?>
                                             <breathless>
      7     All:         ha ha ha
      8     Jane:        and   [so: (.)
      9     Sue:               [<<p>oh NO ;>
      10    Jane:        he came in and said (-)
      11                 f₃: <<h> this WOMan ,        <bright>
```

---

[30] In the absence of content- or style-related hints, attributing every other line to the same figure may be the only way to do so (See iv in the discussion of extract 1). For spoken discourse, this might correspond to the attribution of a new intonation phrase or a more strongly marked prosodic shift to a change of embedded animators.

[31] In the latter case, our data provides some evidence for the prosodic grouping of figures. In the context of the father's complaints in the data Divorced Parents, the two female figures whom he reports as taking his side in the struggle over the child, receive rather similar prosodic designs for the reported speech sequences attributed to them. This design contrasts with that of the distinct 'voice' attributed to his ex-wife.

```
        12                  f₃: I never even bOthered to ask her NAME.
        13                  f₃: it was so biZA(h)R(h)RE ;>
        14                      ha ha ha .h (--) .hh
        15      Sue:            sUre he wasn't aWAKE eh heh
->      16      Jane:           NO .
        17                  f₃: and (-) started QUEStioning me ;> eh heh
        19      Sue:            hm
```

In this episode a story is told about a mutual acquaintance Karen, who stopped Jane's colleague Richard very early one morning on the highway when she saw an 'Apple Valley' sticker on his car. Richard later related the incident to Jane, who now tells it to her friend Sue. The first thing to notice about the way the voicing is handled here is that the first figure to be animated, namely Karen (f_1), is given a breathless voice with high register and allegro speech rate. The next figure, Richard on the highway (f_2), is animated with low register. Both voices are introduced verbally with *and she said* (line 1), *and he said* (line 4) respectively. They have maximally contrastive prosody with respect to register. When now line 6 *do you know Jane Kerr* occurs without verbal projection, its embedded animator can nevertheless be traced due to its characteristic prosodic design: this line has exactly the same register, speech rate and voice quality as do lines 3 and 4, which were explicitly assigned to Karen (f_1).

The scene changes now to a later time, namely when Richard reported the incident to Jane. In this case the relevant figures in the story are Richard the reporter (f_3) and Jane. For the voicing of Richard the reporter, who is introduced verbally in line 10 with *he came in and said*, Jane (the narrator) chooses yet another prosodic format, this time high register in combination with bright voice quality. The next three lines (11-13), with their prosody, are thus attributable to the figure of Richard the reporter (f_3). Line 14 heralds a return to the reporting world: Jane the narrator comments with laughter particles on her own story. Sue provides an evaluation of the story so far in the next turn (line 15). It is Jane's follow-up turn beginning in line 16 which is relevant for the point being made here. Its first component is *no*, delivered with 'normal' prosody. Its second component is a set of two lines which have the prosodic design just associated with the voice of Richard the reporter. Thus we attribute them, and only them, to Richard the reporter (f_3) because they have 'his' prosody. Had *no* been formatted similarly, we would have attributed this component to Richard the reporter as well.

In sum, the attribution of reported speech passages to the appropriate figures is facilitated in spoken discourse by the use of individualized prosodic formats. These individualized prosodic designs may occur with or without a verbal projection of the speech being reported. (It is of course in the absence of a reportative phrase that their contribution to referent tracking becomes crucial.) The prosodic format becomes a 'sign' of its attributed originator. Yet it is

27

important to point out that this 'sign' is not iconic, but rather indexical in nature. This can be seen, for instance, in the following fragment:

(17) *Juliette on Divorce (T12a 1.48.15)*

```
        1       Mia:          and and juliette said;
->      2                  f₁: <<p> yeah,
        3                  f₁: my mom> (--) went out of TOWN to get a
                           f₁:                          divOrce;
        4                  f₁: so that (-) so that my DAD couldn't get
                           f₁:                          NASty
                               <<l,p> or something.>
        5                      (2.0)
        6                      <<n> or wouldn't ´know about it - >
        7                      she made /SOME (.) de/RO:Gatory /CO:Mment;(-)
        8                      about hOw shE had to
                           f₁: (.) SLIP out of town
                               <<all> in order to divOrce you>.
        9       Bill:          .hhh
        10      Mia:           and I(h) ch-
        11                     i said;
->      12                 f₂: (-) <<h,p> JuliETTE>!    <singsong>
->      13                 f₁: <<l,p,len>´SO YEAH;>     <breathy>
        14                 f₁: it was BETter that way;> (--)
        15                     and i tho[ught
        16      All:                    [<<f>ha ha ha>
        17      Mia:       f₂: <<f>Okay that's what YOU think -
        18                 f₂: I don't know though
                               i said>
```

Mia is reporting to her brother Bill a conversation she recently had with his young daughter Juliette. In lines 2-4 she animates Juliette's voice (f₁): *yeah, my mom (--) went out of town to get a divorce; so that (-) so that my dad couldn't get nasty*. The reported speech is verbally projected with *and and Juliette said* (line 1); it has weak prosodic marking with piano volume. Towards the end of line 4 Mia displays some uncertainty about Juliette's exact words with *or something*. Following a two-second pause (line 5) she then reformulates Juliette's comment about her father as *wouldn't know about it* (line 6). But still more work seems required. In lines 7-9 Mia again rephrases Juliette's words, this time in indirect speech: *she made some derogatory comment...* In doing so, however, she shadows Juliette's choice of expression with the prosodically framed *slip out of town*. Next Mia animates the report of her response to Juliette in a voice marked prosodically by high register, piano volume and singsong pitch modulation (f₂, line 12). Finally she animates Juliette's reply (f₁) *so yeah, it was better that way* (lines 13-14) with low register, piano volume, slow speech rate and a breathy voice quality.

Notice that the prosodic design of Juliette's reply (lines 13-14) stands in marked contrast not only to the format used for Mia the confidante (line 12) but also to Juliette's earlier format (lines 2-4). Whereas the soft voice of Juliette's first lines could be said to be mildly iconic of a small child, her later voice is by no means so: it is that of a world-weary grown-up.[32] (In this case the prosodic design has been chosen not only to distinguish one figure from another but also to typify the figures and their characteristic behaviors and attitudes.) Notice too that Mia the confidante's voice also changes the second time it appears: in lines 17-18 it is markedly loud, similar to the laughter (line 16) which it follows, and no longer has the high register or singsong modulation it had in line 12.

Fragment (17) thus demonstrates in two different instances how the prosodic formats selected for figures whose speech is being reported may be altered to suit local needs. In the case of Juliette, the voice is modulated to fit the words which are being put into her mouth. In the case of Mia the confidante, the voice is altered to match surrounding prosody and perhaps thereby suggest collusion. Because it is context-sensitive to this extent, the prosodic formatting of reported speech cannot be thought of as purely iconic signalling; it is also indexical in the sense that it points to the presence of an embedded animator in the co-/context, without necessarily representing that animator as such.

When compared to example (16), fragment (17) is revealing in another sense. Recall that in the latter case different figures were given different prosodic designs and that these individualized patterns were used consistently thereafter. It was this which enabled successful referent tracking in the absence of verbal projectory devices. Fragment (17), however, shows that a figure can be tracked despite the fact that its prosodic design is not consistent. Lines 13-14 are attributable to Juliette although they lack explicit introduction and do not have the same prosodic formatting as her earlier voicing. On what grounds is this inference made? Prosody, we assume, is 'overridden' here by everyday understandings of conversational exchanges and how they are carried out. An appeal such as *Juliette!* is uttered only in second position to some specifiable first and it makes a response addressed to what it is doing expectable in a third turn. When a new intonation phrase now commences with a different prosodic configuration, we are inclined to attribute this to the speaker of the first turn because this is the way such conversational exchanges are typically structured. Just as in written discourse, referent tracking

---

[32] The narrator thus appears to be suggesting a role reversal: Juliette the child becomes adult-like in her perception of the world.

in conversational reported speech may be accomplished by an understanding of the 'logic' of conversational exchange alone.[33]

5. Conclusion

In this paper we have compared the prosodic marking of reported speech in spoken conversational discourse with the typographical marking of direct reported speech in written discourse. Whereas direct reported speech in writing is precisely delimited by quotation marks and speakership is (at least initially) attributed explicitly, any and all forms of reported speech may receive prosodic marking in spoken conversational discourse. Among the prosodic and paralinguistic devices used most frequently are global pitch (register) and loudness shifts, global changes in speech rate and shifts to isochronous timing. If an isochronous rhythm begins with the reportative phrase or before, a rhythmic pause may cue the beginning of the quotation itself. Paralinguistic voice quality effects (e.g. breathiness, nasality, brightness, etc.) often accompany these prosodic shifts.

The prosodic marking of reported speech in spoken discourse is not wholly comparable to typographical marking in writing, because it is a stylistic device rather than a norm: it may be used to signal reported speech or not, depending on speakers' local goals and strategic choices. Moreover, the prosodic marking of reported speech in spoken discourse need not delimit or frame the quoted material as precisely as do quotation marks. Prosodic shifts are sometimes used as flags positioned in the vicinity of the reported speech passage, foreshadowing or echoing a figure's voice with its specific prosodic design. Individualized prosodic designs often serve as indexical devices when several figures are being animated in speech. But even in their absence referent tracking is usually possible — as in written discourse — based on inferences from the structure of 'real' conversational exchanges.

---

[33] We are reminded here of Hanks' (1990) discussion of reference as a communicative activity. He points out that there are two kinds of referential use of definite decriptions: (1) in an anaphoric chain, (ii) not in an anaphoric chain. "The former individuate a referent through a relation of coreference with preceding words in the discourse. The latter usually rely on presupposed background knowledge shared by participants: a kind of tacit anaphoric relation in which prior experience secures uniqueness of reference". Hanks' point is that "...the uniqueness and identifiability of a referential object are not properties of a pure conventional code but of the situated use of a code in a pragmatic frame" (p. 34) - a statement applicable to the use of prosodic designs for referent tracking in reported speech as well.

Transcription Conventions[34]

Sequential structure

[ ] [ ]	overlap
=	quick, immediate connection of new turns or single units

Pauses

(.)	micro-pause
(-), (--), (---)	short, middle or long pauses of 0.25 - 0.75 seconds, up to ca. 1 second
(2.0)	estimated pause of more than 1 second

Other segmental conventions

and=uh	slurring within units
:,: :,:::	lengthening, according to its duration
uh,ah, etc.	hesitation signals, so-called "filled pauses"

Laughter

haha hehe hoho	syllabic laughing
((laughing))	description of laughter

Reception signals

hm, yes, yeah, no	one syllable signals
hm=hm, yea=ah,	two syllable signals

Accents

ACcent	primary or main accent
Accent	secondary accent

Pitch step-up/stepdown

↑	pitch step down on the following syllable
↓	pitch step up on the following syllable

Change of pitch register

<<l>	>	low pitch register
<<h>	>	high pitch register
<<n>	>	use of small segment of speaker's voice range
<<w>>		use of large segment of speaker's voice range

Final pitch movements

?	high rise
,	mid-rise
-	level pitch
;	mid-fall
.	low fall

Volume and tempo changes

<<f>	>	forte, loud
<<ff>	>	fortissimo, very loud

---

[34] Based on GAT, Gesprächsanalytisches Trankriptionssystem (Selting et al. 1998).

<<p>	>	piano, soft
<<pp>	>	pianissimo, very soft
<<all>	>	allegro,fast
<<len>	>	lento, slow

Breathing in and out

.h, .hh, .hhh	breathing in, according to its duration
h, hh, hhh	breathing out, according to its duration

Rhythm

/	/	Isochronous beats; distance between bars indicates length
/	/	of rhythmic units

Other conventions

( )	unintelligible passage, according to its duration
(such)	presumed wording
->	specific line in the transcript which is referred to in the text

References

Auer, P., E. Couper-Kuhlen, F. Mueller (in print). *Language in Time.* New York: Oxford University Press.

Auer, P., & di Luzio, A. (Eds.). (1992). *The Contextualization of Language*. Amsterdam: Benjamins.

Bruenner, G. (1991) Redewiedergabe in Gesprächen. *Deutsche Sprache,* 1-15.

Clark, H. & Gerrig, R. (1990): Quotations and demonstrations*.* In: *Language* 66, 764-805.

Couper-Kuhlen, E. (1993). English Speech Rhythm: Form and function in everyday verbal interaction. Amsterdam: Benjamins.

Couper-Kuhlen, E. (1998). Coherent voicing: On prosody in conversational reported speech. *InLiSt. Interaction and Linguistic Structures*, *1*, 1-26. Fachgruppe Sprachwissenschaft, Universität Konstanz.

Cruttenden, A. (1986). *Intonation.* Cambridge: University Press.

Goffman, E. (1981). *Forms of Talk.* Oxford: Blackwell.

Guenthner, S. (1997a). The contextualization of affect in reported dialogues. In S. Niemeier & R. Dirven (Eds.), *The Language of Emotions. Conceptualization, expression, and theoretical foundation* (pp. 247-276). Amsterdam: Benjamins.

Guenthner, S. (1997b). Direkte und indirekte Rede in Alltagsgespraechen - zur Interaktion von Syntax und Prosodie in der Redewiedergabe. In P. Schlobinski (Ed.), *Zur Syntax des gesprochenen Deutsch*. Opladen: Westdeutscher Verlag.

Guenthner, S. (1997c). Stilisierungsverfahren in der Redewiedergabe - die 'Ueberlagerung von Stimmen' als Mittel der moralischen Verurteilung in Vorwurfsrekonstruktionen. In B. Sandig & M. Selting (Eds.), *Sprech- und Gespraechsstile.* Berlin/New York: de Gruyter.

Gumperz, J. (1982). *Discourse Strategies.* Cambridge: University Press.

Hanks, W.E. (1990). *Referential Practice.* Chicago: Chicago University Press.

Holt, E. (1996). Reporting on Talk: The Use of Direct Reported Speech in Conversation. *Research on Language and Social Interaction 29(3),* 219-245.

Leech, G. N., & Short, M. H. (1981). *Style in Fiction. A linguistic introduction to English fictional prose.* London and New York: Longman.

Mayes, P. (1990). Quotation in Spoken English. *Studies in Language 12-2,* 325-363.

Romaine, S. & Lange, D. (1991). The Use of *Like* as a Marker of Reported Speech and Thought: A case of Grammaticalization in Progress. *American Speech 66,* 227-279

Ruehmann, M. (1999). Reported Speech, 'Voicing', and 'Pre-Voicing': Typification Strategies in Everyday Conversation. M.A. Thesis. Dept. of Linguistics, University of Constance.

Selting, M., Auer, P. et al (1998). Gespraechsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, *173*, 91-122.

Tannen, D. (1989). "Oh talking voice that is so sweet": constructing dialogue in conversation. In: Talking Voices. Cambridge: University Press, 98-133.

Yule, G. (1995). The Paralinguistics of Reference: Representation in Reported Discourse. In: G. Cook/B. Seidlhofer (Eds.) *Principle & Practice in Applied Linguistics.* Oxford: Oxford University Press, 185-196.

Literary source:

Jane Austen: *Pride and Prejudice.* New York: Dell Publishing, 1959.

Previously published in this series: