Manuscript Number:

Title: The coordination of talk and action in the collaborative construction of a multimodal text

Article Type: Full Length Article

Corresponding Author: Dr Rod Gardner,

Corresponding Author's Institution: Griffith University

First Author: Rod Gardner

Order of Authors: Rod Gardner; Mike Levy

Abstract: This paper explores how speech and action are coordinated in a web-based task undertaken by two high school students working collaboratively at the computer. The paper focuses on the coordination involved in the interactions between the two students and the computer screen, keyboard, and mouse, and explores the temporal synchrony and 'matching' points between speaking and typing, and speaking and mouse movements, within and between participants. Examples include coordination of speaking words aloud whilst typing, coordination of reading aloud from the screen and mouse movements, and coordination between participants, as when one individual is typing and the other talking. The discussion draws on the literature describing the coordination of language and action, kinesic behaviour, and nonverbal communication, including gesture, which have the potential to mediate conversation. Results indicate most coordination of talk and action is at the beginning of the activity. Sometimes work is done to ensure coordination, either by slowing down the talk or pausing or stretching sounds mid-utterance. Talk that is coordinated temporally to some action on the screen is precise; in other words even when action and talk are mismatched (e.g., she is not talking about what she is doing), talk and action can start and finish together.

Griffith
UNIVERSITY

Griffith Institute for Educational Research

Social Sciences Building (M10)
Mt Gravatt campus, Griffith University
176 Messines Ridge Road
Mount Gravatt, QLD 4122

Telephone   +61 (0)7 3875 3472
email: r.gardner@griffith.edu.au

Wednesday, March 11, 2009

Professor Jacob Mey
Editor
Journal of Pragmatics

Dear Professor Mey,

Please find enclosed the manuscript for a paper by Mike Levy and me which we are submitting to the Journal of Pragmatics for consideration for publication.

We look forward to hearing from you in due course.

Yours sincerely,

Rod Gardner
Associate Professor
Griffith Institute for Educational Research
Griffith University

Associate Professor Rod Gardner
Griffith University

The coordination of talk and action in the collaborative construction of a multimodal text

Rod Gardner

[a]Griffith Institute for Educational Research

Griffith University

176 Messines Ridge Road,

Mt Gravatt, QLD, 4122

Australia

Email: r.gardner@griffith.edu.au

Telephone: (61 7) 3735 3472

Fax: (61 7) 3735 5991

Mike Levy

[b]Griffith Institute for Educational Research

Griffith University

170 Kessels Road,

Nathan, QLD, 4111

Australia

Email: michael.levy@griffith.edu.au

Telephone: (61 7) 3735 7525

Fax: (61 7) 3735 6766

All correspondence should be addressed to the first author

Abstract

This paper explores how speech and action are coordinated in a web-based task undertaken by two high school students working collaboratively at the computer. The paper focuses on the coordination involved in the interactions between the two students and the computer screen, keyboard, and mouse, and explores the temporal synchrony and 'matching' points between speaking and typing, and speaking and mouse movements, within and between participants. Examples include coordination of speaking words aloud whilst typing, coordination of reading aloud from the screen and mouse movements, and coordination between participants, as when one individual is typing and the other talking. The discussion draws on the literature describing the coordination of language and action, kinesic behaviour, and nonverbal communication, including gesture, which have the potential to mediate conversation. Results indicate most coordination of talk and action is at the beginning of the action. Sometimes work is done to ensure coordination, either by slowing down the talk or pausing or stretching sounds mid-utterance. Talk that is coordinated temporally to some action on the screen is precise; in other words even when action and talk are mismatched (e.g., she is not talking about what she is doing), talk and action can start and finish together.

Keywords: Coordination; Language; Action; Collaboration; Kinetic Behaviour; Nonverbal Communication; Gesture.

1.1 Introduction

This paper seeks to explore how language and action are coordinated in the context of a web-based task undertaken by two high school students working together at the computer. To complete the task the students were required to collect information on the topic of plastic bags, focusing on possible benefits and problems, and then to use this information to create a web page. In this setting, the dynamics of the interactions between the two individuals and the computer are complex. Typically, although not always, the students' attention is directed towards the computer screen, or parts of it (e.g., the toolbar), and one student is usually in control of the keyboard and mouse. The language that it is produced is either written, in the broadest sense of the term, or spoken. Written material entails multimodal text production, in the form of textual or visual material typed or otherwise used or created on the computer (e.g., text, image, animation etc.). Spoken language, sometimes referred to as 'PC talk' (see Uhlirová, 1994), arises alongside the multimodal text production. In this paper, what we are interested in is how the former is aligned to the latter in conditions where an artefact, i.e. the computer, mediates the processes of text creation.

Uhlirová's (1994) paper on the role of the PC as a relevant object in face-to-face communication is helpful in setting the scene. Uhlirová describes the role of the computer in this setting as that of a third, silent partner in the conversation, and explains that the "ever-changing contents of the screen substantially influence the course of the talk" (p. 514). She describes the PC talk as a *semiotic* act consisting of two components: speech communication between the participants, and nonverbal interaction between the participants and the PC through the keyboard and screen. In seeking to understand the nature of the talk generated in these conditions she emphasises the importance of nonverbal interaction. She goes on to provide an excellent characterisation of PC talk:

> ... some shorter passages of talk (several turns long) are realised by the speakers partly **through** the PC, through what is typed and then available on the screen, and only partly by speech responses. The speaker, instead of telling his partner what the next step of his interaction with the PC is, silently types on the keyboard without saying anything. The other partner reacts verbally to this non-verbal turn. In other words, not only the PC's turns are (inevitably, due to the nature of the PC) non-verbal, but also the participants' turns may be (optionally) non-verbal under certain circumstances. Thus, the speech is reduced to a

minimum and the talk is highly economical at the price of partly suppressing its dialogical nature. (p. 525)

Repeatedly, in this paper, Uhlirová emphasises the active role of the PC in mediating the timing and content of the conversational turns. She describes the PC in this role as a sometime "bearer of talk continuity" (p. 526), and a device for generating "cohesive signals" as a result of its working procedures and processes, and the actions it is required to undertake (p. 527).

The particular focus in this paper is on the timing and coordination of the talk with the actions undertaken on the computer. To better provide a foundation on this topic we need to consider more closely what we know about the coordination of language and action, kinetic behaviour, and forms of nonverbal communication, including gesture, that have the potential to mediate conversation.

2.1 Literature review

In broad terms, when individuals work collaboratively to complete a task, the language generated relates directly to action, as dictated by the task at hand. Normally, within a group task, the actions that are taken need to be negotiated and agreed upon before they are instigated. As Glenberg and Kaschak (2002: 564) claimed, "language almost certainly arose to facilitate coordination of action." Language and action are related dynamically, that is one relates closely to, or is coordinated with, the other in terms of time and place. Goodwin (2000:1489) stressed the importance of looking at "both the details of language use and the way in which the social, cultural, material and sequential structure of the environment where action occurs figure into its organisation." Contextual factors are also critical. Goodwin (2000: 1519) continues:

> … context is not simply a set of features presupposed or invoked by a strip of talk, but is itself a dynamic, temporally unfolding process accomplished through the ongoing rearrangement of structures in the talk, participants' bodies, relevant artefacts, spaces, and features of the material surround that are the focus of the participants' scrutiny. Crucial to this process is the way in which the detailed structure of talk, as articulated through sequential organisation, provides for the continuous updating and rearrangement of contexts for the production and interpretation of action.

This perspective should be brought to bear on the pair of students completing a web-based task at the computer in a high school setting. Language is used to coordinate action, but it is also produced and manipulated by the computer keyboard and mouse. The context is a highly dynamic one where the talk moves forward alongside the ever-changing multimodal text that is being created on the computer. The computer screen, or segments of it, provides a dynamic field of view to which the participants react moment by moment, but it is not always through PC talk, or direct interaction between the participants.

Recalling the brief discussions in the introduction, it appears that the PC may somehow mediate or moderate the talk and the turn taking between the two participants almost like a third person, even though, of course, the computer does not speak. In other words, collaborative work at the computer invokes somehow a form of nonverbal communication or behaviour (NVB). In considering this aspect, Burgoon (1994: 231) describes a "shared coding system." The notion of a socially shared coding system and recognizable interpretations are relevant in this study concerning pair work at the computer in a classroom in a high school, a setting that is highly familiar to students, with its established norms and practices.

NVBs can also constitute a message (Burgoon, 1994: 231). Chamberlin Quinlisk (2008: 31) argues that sometimes these unspoken messages can "outweigh what has been said and change the dynamics of a conversation." In relation to the task-based pair work activity described here, it should not be forgotten that both participants are continuously scrutinising what is being produced on the computer screen; inevitably, some talk will emerge as a result of the negotiated action that is required, choices that are made, and errors that occur in the process. But the point is that the talk derives not only as a direct result of events on-screen, but also as a consequence of more implicitly determined expectations derived from the working context.

In reflecting upon language and action in relation to work at the computer, we can clearly see that our students are creating new materials via the keyboard and mouse through the use of their hands. Such activity relates to body movement or kinetic behaviour that also has the potential to play a significant role in communication (Kellermann, 1992). Kinetic behaviour encompasses all movements of the body, both muscular and skeletal. Kellermann (1992: 240) goes on to state that the "production and perception of these movements are, by and large, out of awareness." We could perhaps apply this to an individual's keyboarding skills, or their facility with the computer mouse,

when an individual's attention and focus is on the task at hand, not the accuracy or the quality of their typing. Kinetic behaviour also has a regulatory function in interaction whereby body movements help maintain and coordinate the flow of speech and turn taking (Duncan, 1972; Gosling, 1981). Gosling (1981) gives examples of how turn-taking behaviours are supported by eye contact, direction of gaze, gestures of the hand, and arm and posture. Importantly, for the present paper, Kellermann (1992: 241) goes on to argue that all body movement is made in time with the rhythm of speech and, more specifically, "stressed tonal nuclei coincide with points of emphasis in kinesic activity".

One example of kinesic activity, of course, is the gesture. Of the many gestures available, the beat gesture has special relevance in light of the discussion so far. Beat gestures are small, rhythmic baton like movements of the finger, hand, or arm that can vary in sense according to the context. A short and single beat can mark an important point in a conversation, whilst repeated beats can hammer home a critical concept (McNeill, 1985). They often serve a pragmatic function, occurring with comments on one's own linguistic contribution, speech repairs, and reported speech. An example is waving one's hand briefly up-and-down along with stressed words in the phrase "Go-aHEAD". McNeill (1985: 359) describes them as a kind of "abstract visual indicator" and emphasises their importance as an "inter-item relation" that often occurs at points of significant discontinuity in the discourse structure. Chamberlin Quinlisk (2008: 31) adds that beats may also regulate the flow of the conversation and position one speaker as more powerful than the other. Also of note with regard to gesture, and with clear parallels with hand movements in relation to the computer mouse, is the notion of rest points either side of periods of motion. Choi and Lantolf (2008) elaborate upon McNeill's (1992) work and describe the stroke of the gesture, its most important phase, as having a semantic and a kinetic component:

> The semantic component depicts the meaning in focus, and the kinetic component, which synchronises with co-expressive speech, is the phase of the gesture during which the most salient movement occurs. The stroke is preceded by a preparation phase, in which the hands move from a rest position to the position where the stroke is executed, and is followed by the retraction phase, in which the hands return to the rest position. (Choi & Lantolf, 2008: 198)

More broadly, Gullberg and McCafferty (2008: 136) explain how gestures are implicated in "interactional work like turn and floor regulation, feedback elicitation, agreement marking, and attention directing via pointing… and also in interactional synchrony or mirroring, with further

ramifications for experiences of sympathy and rapport." While we must be cautious at this early stage in extending any of the implications from the use of the hands for gestures in relation to speech to the use of the hands for keyboarding or the use of the mouse in relation to speech, the potential implications are tantalising.

Importantly, Gullberg and McCafferty (2008) also note the 'Fine-grained coordination of movements both within and between speakers engaged in speaking…" (Gullberg & McCafferty, 2008: 134; see also Birdwhistell, 1970; Kendon, 1972). It is worth emphasising that the coordination is both *within* and *between* speakers. Furuyama (2002) picks up on this point and discusses 'collaborative gesture,' separating between-person (interpersonal) coordination of speech and gesture, a new phenomena, from within-person (intrapersonal) coordination of speech and gesture, the latter being the conventional interpretation. He emphasises in his conclusion the synchronous coordination of more than one meaningfully interrelated stream of actions (e.g., speech and gestures) as not only found within an individual, but also between individuals.

Although there is insufficient space to discuss the evidence here, it is also worth noting that from recent work on the mechanisms of the brain and neural evidence, "language and action recruit overlapping parts of the brain and information coming from both domains can be used in qualitatively similar ways" (Willems & Hagoort, 2007: 286; see Pulvermüller, 2005). Interestingly, also Kelly, Manning, and Rodak (2008) record that "several studies have found that brain regions that process speech also process actions made with the hand" and, moreover, in terms of future research, they advise investigating, "to what extent are hand gestures processed similarly or differently than other actions made with the hand…" (p. 582).

Overall, numerous authors (e.g., Chamberlin Quinlisk, 2008; Goodwin, 1981; Kopp & Bergmann, 2007; McNeill, 1992; McNeill and Duncan, 2000) emphasise the synchronous nature of gestures and speech, or the ways in which they align with one another. Further possible parallels between hand gestures and manipulation of the keyboard and mouse may generate insights. For example, Chamberlin Quinlisk (2008) notes how gestures (substitute keyboarding and use of the mouse) mark discontinuities in discourse and emphasise continuity. Gesture and speech are orchestrated together to form coherent multimedia messages (De Ruiter, 2007), whereas mismatches between gesture and speech predict cognitive transitions (Namy & Newcombe, 2008). Finally, Vygotsky called gestures the 'material carriers' of thinking (Vygotsky, 1987, as cited in McNeill, 2005), and

this seems a particularly apt metaphor for the individual who is reasonably proficient with the keyboard and mouse when working at the computer.

3.1 The study

The case study described in this paper is drawn from a larger, longitudinal Australian Research Council funded study conducted by Griffith University from 2003 to 2007, entitled *Using and creating knowledge in the high school years: Performance, production, process and value-adding in electronic curricular literacy*. This study examined the processes undertaken by secondary school students when creating multimodal products such as *PowerPoint* presentations and websites. The goal was to track the development of the participants' multimodal literacies over a 2-year period. In all, 14 secondary schools participated in this study and the researchers were able to track 115 students from Years 8 and 10 (2004) to Years 10 and 12 (2006). The complete data set consists of survey data, product data, and process data. Process data were collected through recording on-screen activity with video screen-capture software (Camtasia Studio, 2004) as the students collaborated on the online task.

The authors of the current paper determined the student pair for in-depth analysis through an incremental, cyclical process of elimination. Of the students whom we were able to follow successfully from Year 8 to Year 10, we initially chose to focus on a small group represented by one school. By focusing on one school initially, we were better able to isolate differences between the years while keeping differences between schools extant to this equation. Here the year 10 students produced the richer data set (process data). Then, from this group, we selected the students who fulfilled the following three criteria: they had worked collaboratively in pairs, they had fully completed the task, and the video screen data recording were intact and complete. From there, we listened repeatedly to the interactions and sought high quality data that reflected close coordination between the talk and the use of computer. Extended segments of talk were then transcribed for closer analysis. Other factors such as the sound quality of the recording and voices that were distinct and sufficiently different also played a part. Finally, we were in a position to select two students, Anna and Suzanne, as our sample for Year 10 (School 16 - 2004).

The online tasks that students were required to complete were developed in collaboration with year level teachers for both years. Both the 2004 and 2006 tasks were devised as cross-curricular, inquiry-based activities. Of relevance here is the 2004 task, which required students to evaluate possible solutions in response to the environmental threats posed by plastic bags. The task was divided into three phases of development: *researching, designing,* and *reflecting*. Part of the research process required the student to complete two templates: a *concept map* to organise information acquired from given web sources; and a *decision-making matrix* to facilitate transformation of the ideas in their concept map into a preferred solution, as required for their multimodal text. Following these preparatory stages, the students could create a multimodal text using a word processor (e.g., *Word*), presentation tool (e.g., *PowerPoint*), or a web-publishing program (e.g., *FrontPage*). For their project, Anna and Suzanne created a website using *FrontPage* (MicroSoft, 2000). Three hours were requested for completion of the 2004 task, either in one session or successive lessons.

3.1.2 Analysis

In this section, we present three extracts to illustrate the talk-hand-screen coordination in the collaborative task of the two students constructing a web page. In the first of these, one student is using the keyboard and speaks the words as she is typing. In the second, the students are formatting the title page of their web page. In the third, they collaboratively correct a heading line on the page, in the interactionally most complex of the three sequences.

Extract 1

The first extract illustrates one type of word-screen coordination: speaking out the words as one types. In this passage of talk, Anna and Suzanne are revising the third subheading of the first web page in their project. Anna is the student using the keyboard.

Extract 1a:3.14a:717

```
1   Anna:        And then up here, >where (th' part yih) say< Plastic
2                Ba:gs, you sa:y
3                (4.6)
                 ((deletes 'Plastic bags' from webpage))
4   Anna:        <M-m:os:t (1.4) do:n't: (1.8) do:n't: (0.6) realis:e¿
5                (1.8) the ex::te:nt (0.2) of: (0.4) thee (1.8) pla:s::tic-h
6                (1.2) ba::g (2.8) °pollu:tion (1.1) i:n°> (3.9)
                 ((speaking whilst typing))
7   Suzanne:     Australia
8                (4.3)
```

This extract begins, in lines 1-2, with Anna drawing attention to the subheading, 'Plastic bags', as it currently stands on the third line of the web page. This subheading is the same as the first heading on the page, perhaps a reason for changing it. Anna invites Suzanne to look 'up here', near the top of the screen in the centre. She frames the projected new title with 'you say'. In the 4.4 second silence in line 3, she is deleting the existing subheading from the screen before she begins to retype it. As she does so, she reads out what she is typing. Each typed word except the second ('people') is spoken; one of them ('don't') is spoken twice.

Anna orients her talk and typing to occur nearly simultaneously (extract 1b). In extract 1b, an extra line for the keyboard is added. Each stroke of the keyboard is audible, so the letters in the keyboard line represent keystrokes, which are lined up with the spoken words in the line above. Deletion of characters is represented by the character with strikethrough.

A first point to note is that, because it takes Anna longer to type than to say the word, there is a stretching of each word to achieve a temporal mapping onto the typing of the word. In fact, she begins to say most of the words either just before (up to 0.2 seconds) or exactly where the typing begins. In no cases does she begin to speak *after* she has begun to type the word. This is an example of *intra*personal coordination, though by speaking the words out loud, she is making what she is doing available to Suzanne not only visually through the computer screen, but also audibly through her speaking.

Extract 1b:3.14a:717

```
1  Anna:      >An' then up he:re, >where (th' part yih) say< Plastic
2             Ba:gs, you sa:y,


3             |-------------(3.0)-------------|
   Anna:
   Mouse      ((moves to end of 'Plastic bags'))


4             |----------(1.1)---------|
   Anna:
   Keyboard   |s g a b _ e i t s a l P |
              ((deletes 'Plastic bags'))


5  Anna:      |--(0.5)--|<M:o:s:::'|--(-1.0)---|(0.2)|(0.2)|d:o:n::'|t:
   Keyboard   |M  o  s  | t _ p  |e o p l e _|     |d o | e       |nt


6  Anna:      |(0.2)| (0.3) |--(1.3)--|do:n't:h|(0.6)| reali::s:e|-(0.7)--|
   Keyboard         | t n e |         |n' t _|   r| ea li c  | e
|


7  Anna:      |-----(1.1)-----| thee: |e:x:::ten:t:|(0.2)| o:f:
   Keyboard   |s  e _ t h e | _   |e x t e n d| _  | o


8  Anna:      |(0.4)| thee:|---(1.8)---| pla:s::|(0.3)|ti:c::|-(0.4)-|
   Keyboard   |f _ t| h e | _           | pl  |a s| t  | i c _ |


9  Anna:      |--(0.8)--| ba:::g|-(1.0)|---(1.8)---|°p'llu:t:ion:|
   Keyboard          b |   a g|   _|      p o| l l  u tion |


10 Anna:      |-(1.1)--| in::°> |-----(2.3)----|
   Keyboard   |_ i    n|       |_ a s u u s  |


11 Anna:      [--- (0.7)---]----(0.9)----|
   Keyboard   [ a   [A u ]s t r a l i a|


12 Suzanne:   [Austra[lia,]


13            (0.9)


14 Anna:
   Keyboard   .
```

There are, however, some mismatches between speaking and typing, either at the beginning or the end of the word, so that the production of this sentence does not proceed entirely smoothly. These mismatches are mostly associated with corrections to the typing, leading to longer silences between words: There is a 1.8 second silence between the two sayings of 'don't' (lines 5-6 of extract 1b), 1.8 seconds between 'realise' and 'the' (lines 6-7), and 3.9 seconds between 'in' and Suzanne's saying of 'Australia' (lines 10-12). The time needed to execute these corrections explains these longer silences. The last of these silences, at the end of the typing of the sentence, is particularly long, and this length is partly explained by the fact that the repair is a more complex one ('asu' repaired to 'Aus'). It is striking that Suzanne comes in on the last word, and not before, suggesting that her tolerance for the silence has reached a 'maximum' (cf. Jefferson's (1989) standard maximum silence of around 1 second for ordinary conversation, rather than this task-based talk).

There is evidence of three different ways of correcting: immediate correction of a single grapheme; corrections of multiple errors carried out after complete typing of a word; and correction whilst reviewing the sentence after it has been fully typed. The first, immediate correction, occurs in line 6 of extract 1b, where Anna replaces the incorrect letter 'c' with the correct one 's' immediately after she has typed it (a phonetic-graphemic correction). The second kind of immediate correction is where, after typing a whole word, the 'don't' in lines 5-6, she deletes the erroneous letters and then types them correctly. There are in fact two errors here: Anna had inserted an 'e' after the 'o' (cf. 'does') and also omitted the apostrophe; this is a form of a lexical item, rather than single grapheme, correction. It is noteworthy that with 'realise', she corrected the letter immediately it was typed, which gave her no time to repeat the word. With 'don't' she had to return deeper into the word, and make two corrections, giving her the time to repeat the word aloud. The correction of 'Australia' at the end of the sentence occurs in a similar manner to 'don't', except that it is accomplished collaboratively. After Anna types 'asu', she corrects the two errors by deleting the 'u' and 's', and then pausing momentarily, then capitalising the 'a' and reordering the 's' and the 'u'. During this pause, Suzanne comes in for the first time, to say 'Australia', just at the point at which Anna capitalises the 'a', that is, at the point where she begins to type the word for the second time, demonstrating *inter*personal coordination.

The third type of correction (and the fourth correction overall), in contrast to the ones above, is delayed. Initially, the word 'extent' is entered incorrectly as 'extend'. Unlike the first two, where the correction was temporally close to the error, 'extend' was not corrected until

Figure 1

Plastic Bags

Their impact on the environment

Most people doent

Heading read aloud
whilst being typed
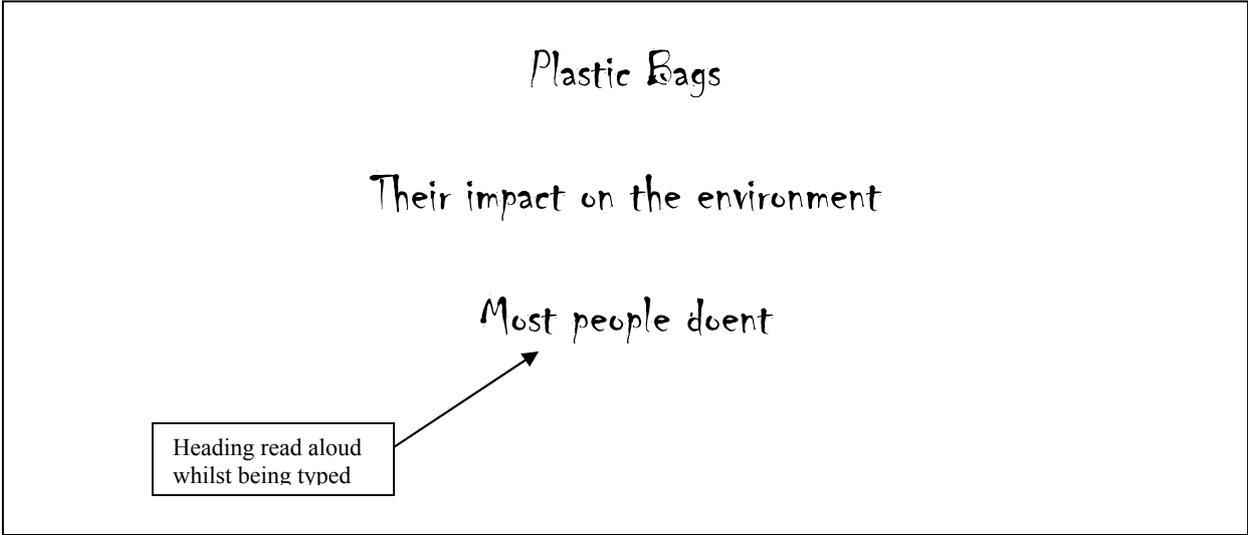
Retyping the first heading

the whole sentence was being reviewed after she had finished typing it. Throughout the typing to this point, the on-screen video data reveal that the mouse was in a static position to the right of the screen. After completing 'Australia', Anna moves the mouse to the left end of the sentence, and guides the cursor (visible in Camtasia as a yellow circle) over the words from left to right. When she reaches 'extend', the mouse is brought to rest and clicked just to the right of the word 'extend', which is corrected. It thus appears that the mouse and eyes scanned the words of the sentence together. Note also that all four corrections were made without comment or negotiation between the partners, and most is done intrapersonally, but made visible and mostly audible to the partner.

Extract 2

In extract 2, Anna and Suzanne are working on the appearance of the page. This spate of talk occurs shortly after extract 1, and the focus is now on the list of four dot points in the middle of the page (see figure 2). Anna states that she is going to increase the font size, and change the font, which is currently the 'default' font for *FrontPage*. It is notable that, despite the mismatch between what she is beginning to say (about increasing the font size) and what she is actually doing (bolding the first dot point), the start and finish of the speaking of this utterance ('and then make this all') and the action of bolding the dot point on the screen are perfectly synchronised (see extract 2b below). Hand movement (mouse) and speech (the utterance) are thus precisely coordinated.

Extract 2a:3.14a:765:16'05"

```
1                (0.6)
2  Anna:         And then make this a:ll
3                (4.8)
4  Anna:         th'n I'll make it a bit bigger;
5  Suzanne:      Yea:h.
6                (9.5)
7  Anna:         >□I's think'n y'know< we should change the font for
8                this pa:ge.
9                (1.2)
10 Anna:         >Like we only< just keep the- (0.2) title in Chiller?
11               (0.6)
12 Suzanne:      Yeap
13               (3.4)
```

There are three distinct activities in this short extract. Whilst Anna is proposing, in line 2 of extract 2a, to increase the font size of the dot points (to 14pt), she is actually engaged in an prior activity, namely fixing an earlier oversight, which had left the first dot point on the page, '7 billion used a year', in plain text (see figure 2), which she now bolds. So it could be said that she completes the first activity without comment and in 'overlap' with the second. Whilst Anna is speaking line 2, she drags the mouse over the sentence from right to left to highlight it (line 2 of extract 2b, indicated by the left-pointing outline arrows in the 'mouse' line. See the glossary for explanation of the symbols for mouse movements).

During the 4.8 seconds' silence in line 3, Anna completes the initial activity of bolding the first dot point. She moves the mouse to the toolbar to click on 'bold', then moves the mouse down the screen to remove the highlighting by clicking on the screen. She now embarks on the second activity in the extract of increasing the font size of all four dot points. She moves the mouse down the page to just beyond the bottom right corner of the four dot points. At precisely the point when she reaches this position with her mouse, she begins to say 'then I'll make it a bit bigger', which is a resaying (with a slight change in wording) of the utterance begun, and abandoned, in line 2. Thus, the action of highlighting to prepare to increase the font size begins at precisely the same time as Anna starts to speak line 4. There is, however, no matching of the end of the utterance and the completion of the highlighting, as the action

Figure 2

Plain text to be
converted to bold

Plastic Bags

Their impact on the environment

Most people don't realise the extent of the plastic bag pollution in Australia.

For example:

- 7 billion used a year
- **Less than 3% of total plastic bags are taken to the central collection point, other than a dump, for recycling.**
- **Plastic bags are produced from polymers derived from petroleum. The amount of petroleum used to make a plastic bag would drive a car about 11 metres.**
- **230,000 supermarket bags are dumped in landfills every hour.**

Four dot points

takes 0.8 seconds longer to be completed than for Anna to make her announcement (line 4) and Suzanne's response (line 5).

During the 9.5 seconds' silence that follows in line 6, Anna moves the cursor to the pull-down menu on the toolbar to select a bigger font size. She moves across to the pull-down menu to select a different font and, on opening this menu, she makes her next suggestion, in lines 7 and 8, to change the font, which is currently the default font.

Extract 2b:3.14a:765:16'05"

```
1                   (0.6)
2  Anna:      And then make this a:ll
   Mouse      ⬆⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬆
3a                  |----(1.7)----|
   Mouse      ⬆↗↗↗↗↗↘BOLD⬆
3b                  |--------(1.1)--------|
   Mouse      ⬆↙↙↙↙↘DE-HIGHLIGHT⬆
3c                  |----(2.0)----|
   Mouse      ⬆⬅↖↘⬆↗↘↓↙↙⬆
4  Anna:      th'n I'll make it a bit bigger;
   Mouse      ⬆↓↓↓↓↓↓↓↓↓⬅⬅⬅⬅⬅⬅⬅⬅⬅⬆=
5  Suzanne:   =Yea:h.
              =⬆⬅⬅⬆=
6a                  |---(0.8)---|
   Mouse      ⬆⬅⬅⬅⬅⬅⬆
6b                  =|-----------(4.4)-------------|
   Mouse      =⬆↗↗↗↗↗↘PDM↓↓↓↓↘FONT SIZE 14⬆
6c                  |----------(4.3)-----------|
   Mouse      ⬆↘→↗↗↑↖↖⬅⬅↘PDM-FONT→→⬆
7  Anna:      >□I's think'n y'know< we should change the font for
   Mouse      ⬆⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅⬅↘CLOSE PDM FONT⬆
8             this pa:ge.
9                  |---(1.2)---|
   Mouse      ⬆↗↗↗↗↗↗↗⬆
10 Anna:      >Like we only< just keep the- (0.2) title in Chiller?
   Keyboard        C H  i  l l   e      r        ↘ENTER
11                 (0.6)
12 Suzanne:   Yeap
13                 (3.4)
```

In the remaining 8.7 seconds of silence (line 6), two further actions occur after the highlighting has been completed: Anna moves the mouse to the font size menu and selects size 14 font (taking about 4.4 seconds), and the activity of increasing the font size is now completed. Now Ann moves across the toolbar to the font selection pull-down menu, which she opens (which takes another 4.3 seconds). This occurs without talk. Anna has in fact unilaterally embarked upon the third activity, changing the font type. This pull-down menu is open, and she has moved the pointer to menu scrollbar. Next, instead of scrolling to locate the font, she abandons the scrollbar, and takes the shortcut of typing the name of the font directly into the font name window. All this occurs before she speaks line 7.

In lines 7 to 8, Anna articulates her proposal to the change of font type. Two significant coordination points are evident during this utterance. Remember that fonts can be selected either by scrolling down to the pull-down menu, or by typing in the name of the font in the small window with the menu closed. Just at the point when she starts to move the mouse from the scrollbar on the open pull-down menu to the font type window, she starts to speak line 7. Whilst she is saying these lines, she moves to the window and closes the pull-down menu. At the final word of her utterance, 'page' (line 8), which is stretched, the pointer comes to rest on the font window, ready to type in the font name. This is another example of almost perfect coordination of speaking and action. It is intraperson coordination, in which what she is doing (her action) is precisely coordinated with what she is saying.

In the silence of 1.2 seconds that follows, Anna moves the mouse to a resting point towards the top right. She next begins to type in the name of the font, Chiller, in the font selection window. A further example of coordination of talk and typing is evident here. Whilst saying, 'Like we only just keep the title in Chiller' (line 10), she types in the font name. There is a very slight time lag between starting this utterance and typing 'Chiller' (cf. very much like what happened with the slight foreshadowing of typing by the talk in extract 1). What is emphatic about this utterance is its ending, with the word 'Chiller', which coincides perfectly with the audible hitting of the 'enter' key. This, however, requires some work for Anna. Just after she says 'Like we only', she starts to type 'Chiller'. In order to maintain her talk-action coordination to the end of the utterance, she has to pause briefly in the middle of the utterance: the 0.2 seconds' silence after the word 'the'. This allows for perfect coordination, with the word 'title' coinciding with the typing of the 'r' in 'Chiller', and the saying of

'Chiller' then coinciding with the hitting of the 'enter' key, which is the point at which the font changes on the page.

What is perhaps even more surprising, given the points of coordination just described, is that Anna is signalling that she is not satisfied with Chiller as the font for the whole page. She has already made up her mind that another font is needed. This decision plays out in the ensuing discussion, where ultimately the font 'Elephant' is chosen for the bullet point text in focus here. She is saying she wants to change the font *from* Chiller, even as she is changing it *to* Chiller.

Extract 3

The third extract is from talk that occurs a few minutes earlier than the two extracts above. The two students are at the very beginning of their work on constructing the web page, typing in the first heading. At this point, Anna is working off the keyboard, and Suzanne is working with the mouse. In line 222, about 30 seconds before the extract begins, Anna says to Suzanne, 'Just click X', which tells us that Suzanne is in control of the mouse. In addition, from line 255 (line 13 of extract 3), both cursor and the mouse move simultaneously, which is very unlikely to occur if only one of them had control of both keyboard and mouse. Furthermore, what they say during the extract concurs with Anna typing and Suzanne moving the mouse.

The bare transcription of words (extract 3a) is opaque. Line 1 is incomplete, as is line 3. Line 5 is very quiet (probably 'I dunno'), and then there is heavy deixis in line 7 (what is to be fixed up is not clear). There is a go-ahead (line 9), and a 'response cry' 'Oops' (line 11), an indication that 'spelling' is the problem (line 13), and then Suzanne says, 'Their impact on the environment' (line 16), apparently out of the blue. Only Anna's last utterance in lines 17 and 18 makes specific and clear reference to what they are doing.

Extract 3a:3.7:243:5'16"

```
1  Anna:       U::m;=and then you put (1.0) ↑plastic-
2              (5.0)
3  Suzanne:    I don't think
4              (1.0)
```

```
5    Suzanne:    >°(I dunno)°<
6                (8.5)
7    Suzanne:    Can I fix: something up?
8                (0.8)
9    Anna:       Go ahea-
10               (1.0)
11   Anna:       Oops::
12               (0.5)
13   Anna:       Spell↑ing:, .h=
14   Suzanne:    =Yea:h,
15               (6.0)
16   Suzanne:    °·hhh Their ↑impact-↑ (.) on the environment°
17   Anna:       ↑Let's have a↑ different background this time,=
18               let's have l[ike a naturish sort of background.
19   Suzanne:                 [Yeah
20               (6.0)
```
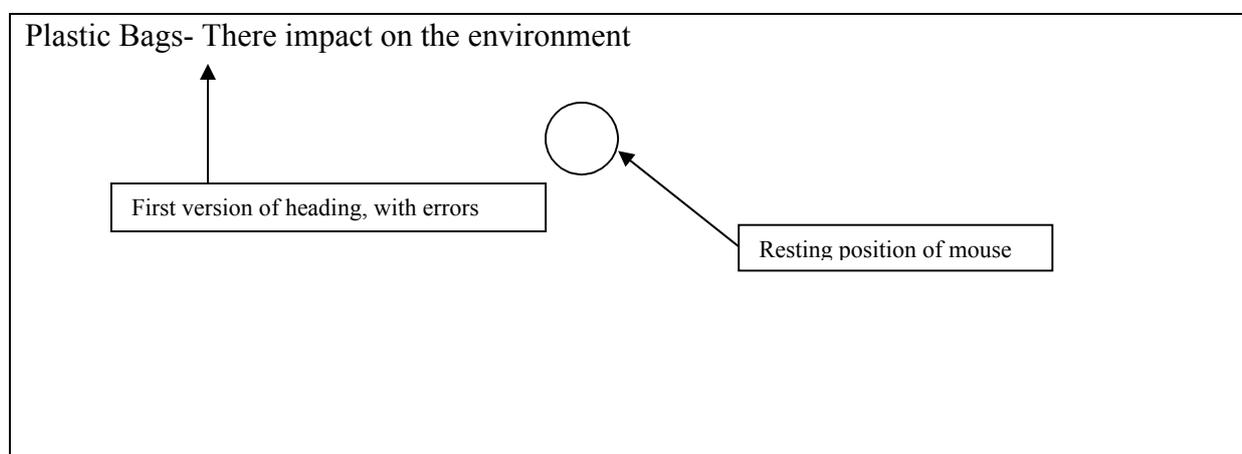
When the video of what is happening on the screen is scrutinized, what they are talking about becomes much clearer (as captured in extract 3b). The two girls are starting to type the first page of their web page. There is an incomplete utterance in line 1, 'Um, and then you put (1.0) plastic'. The saying of the word 'plastic' is designed to accompany its typing. Just over 1 second after 'and then you put', Anna begins to utter the word 'plastic', which in turn occurs 0.2 seconds before she begins to type the word. This is very similar to the shadowing of typing and speaking seen in extract 1. This time, however, she stops speaking after the first word, and the sentence is completed by typing alone. She takes about 16 seconds to type the heading, 'Plastic Bags- There impact on the environment' (see figure 3a).

Extract 3b:3.7:243:5'16"

```
1    Anna:        U::m;=and then you put (1.0) ↑pla|stic-
     Keyboard (H)                                  |p  l
2b                |------------------------(5.0)-----------------------|
     Keyboard (H)   a (0.6) s t i c (0.4) _ b a g s _ (0.6) _ s̶ g̶ a̶ (.) b̶=
3    Suzanne:     =I: don't ↑think,
4                 |-(1.0)-|
     Keyboard (H)  B a (0.4)
5    Suzanne:     >°(I dunno)°<
6a                |----------------(3.0)----------------|
```

```
     Keyboard (H)   g s (0.3) - _ (0.3)  T h e r (.) e (0.6) _
6b                  |-----------------------(5.5)-----------------------|
     Keyboard (H)   i m (.) p a c t _ o n _ t h e _ e n (.) v i r (.) o n m
7  Suzanne:         [Can I fix:] something up?
     Keyboard (H)   [  e    n    t ]
8                   (0.8)
9  Anna:            Go ahea-
10                  (1.0)
11 Anna:            Oops:[:
     Cursor (H)          [⇕
12                  |--(0.5)--|
     Cursor (H)     ⇕ ⇕ ⇕ ⇕ ⇕
13 Anna:            [Spell↑ing:,  .h]=
     Cursor (H)     [⇕ ⇕ ⇕ ⇕ ⇕ ⇕ ⇕ ⇕]=
     Mouse   (N)    [↖ ↖ ↖ ↖ ↖ ↖ ↖]=
14 Suzanne:         =[Yea:h, ]
     Cursor (H)      [⇕ ⇕ ⇕ ⇕]
     Mouse   (N)     [↖ ↖ ↖ ↖]
15a                 |-------------(5.3)--------------|
     Keyboard (N)   ⌐↰  ↗ ≠ (1.3) e / (0.5) ≠ e (0.7)
16 Suzanne:         °·hhh Their ↑impact-↑ on the environment°=
     Mouse          ↰↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↰⇦⇦⇦⇦⇦⇦⇦⇦⇦↰
17 Anna:            =↑Let's have a↑ different background this time,=
     Mouse          ↰↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↘ ↰⇦
18                  let's have l[ike a] naturish sort of background.
     Mouse          ↰PDM-FONT       ⌐PDM-FS ↓ ↓ ↓ ↓ ↓ ↓ ⌐FS24↰
19 Suzanne:                   [Yeah ]
20                  (6.0)
```

Whilst Anna says nothing during these 16 seconds, she does make two mistakes, one of which is repaired almost immediately, the second of which she attends to only after she has finished typing the whole sentence. Suzanne's 'I don't think' and '(I dunno)', lines 3 and 5, remain opaque to us, but it is striking that they occur just after Anna begins repairing the first error, namely typing the 'b' of 'bags' in lower case. Furthermore, both of Suzanne's utterances are placed in pauses in the typing around the repair of this error. Thus, Suzanne's utterances appear to be oriented to trouble with the typing, as she remains silent during the subsequent error-free typing.

However, it appears that Suzanne has in fact noticed the second error: 'there' for 'their'. She withholds from comment until just before Anna finishes typing. Whilst Anna is typing the last three letters of the heading, the 'ent' of 'environment', Suzanne asks if she can 'fix something up'. She has thus correctly predicted where the typing of the heading will come to completion, and coordinated the start of her proposal to coincide with a point just before the completion of Anna's typing – an interpersonal coordination. What Suzanne wants to fix up is, first, 'their' for 'there'; perhaps, however, she is also considering the appearance of the title, which is the task upon which they next embark.

Figure 3a

Plastic Bags- There impact on the environment

First version of heading, with errors

Resting position of mouse

First attempt at typing the heading

A second after giving the go-ahead to Suzanne to 'fix something up', Anna's 'oops' (line 11) suggests she has noticed that something is wrong. She, remember, is the one using the keyboard, and she begins to use the back arrows to move from the end of the heading towards the trouble source. She reaches the space between the 'n' and 'v' of 'environment', still four words away from 'there'. A short while before she reaches that point, when the cursor bar reaches the space between 'n' and 'm' of 'environment, Suzanne starts to move the mouse from its resting place below of the 'ment' of 'environment' (figure 3a) towards the top left to trouble source word, 'there', where the yellow circle of the mouse comes to rest (figure 3b). At this point, both the mouse and the keyboard cursor are moving at the same time: Anna is using the keyboard, whilst Suzanne is using the mouse.

Suzanne reaches the word 'there' as Anna's cursor has reached the space between 'n' and 'v' of 'environment'. Suzanne now clicks on the space between the 'r' and the second 'e' of 'there', and the cursor jumps to that spot. Suzanne has, in effect, snatched control of the cursor bar. She then moves the mouse towards the top right to a resting spot above the 'on'. It seems that Suzanne now takes over the keyboard to correct 'there' to 'their', taking up Anna's granting of her request to 'fix something up': it is Suzanne who reads the now corrected heading, 'Plastic Bags- Their impact on the environment' in line 16.

Figure 3b
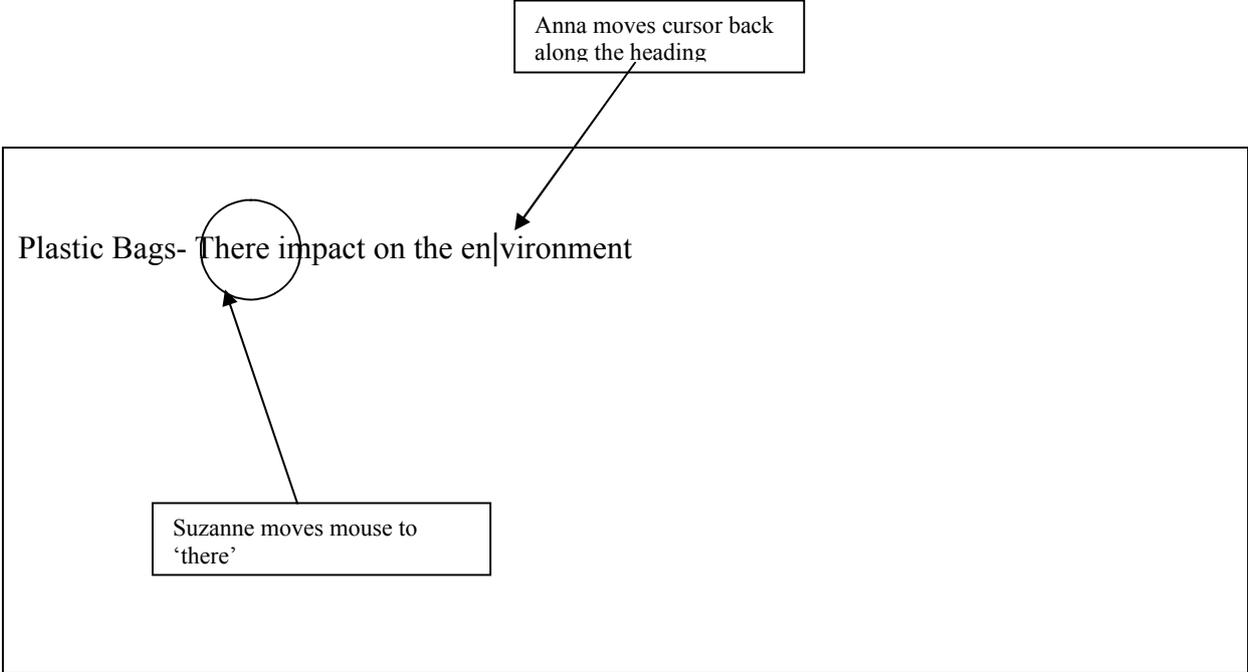
Correcting the heading

As she is reading the heading aloud (line 16), Suzanne moves the mouse from right to left across the heading to highlight it (see figure 3c). She completes her reading of the title just as the highlighting appears on the screen, a further example of intrapersonal coordination between talk and mouse movements. Latching on to this utterance, Anna proposes a new activity, which is to have a different background for the page (lines 17 to 18), which Suzanne agrees to with her 'Yeah' in line 19. They are now out of alignment in terms of their actions: Suzanne still has to complete her task of changing the appearance of the heading, whilst Anna

is proposing they change the background. Nevertheless, Suzanne agrees to this proposal, whilst engaged in the different activity.

Whilst Anna is saying, 'Let's have a different background this time, let's have a naturish sort of background' (lines 17 to 18), Suzanne moves the mouse to the font size pull-down menu, scrolls down to 24pt, and just as Anna completes the second 'background' (the end of line 18), Suzanne clicks on '24pt', and the size of the heading jumps to the larger size. After this, Suzanne goes on in the ensuing silence to underline the heading and to change the font.

Figure 3c

Plastic Bags- Their impact on the environment

Final position of mouse after highlighting

Suzanne highlights the heading
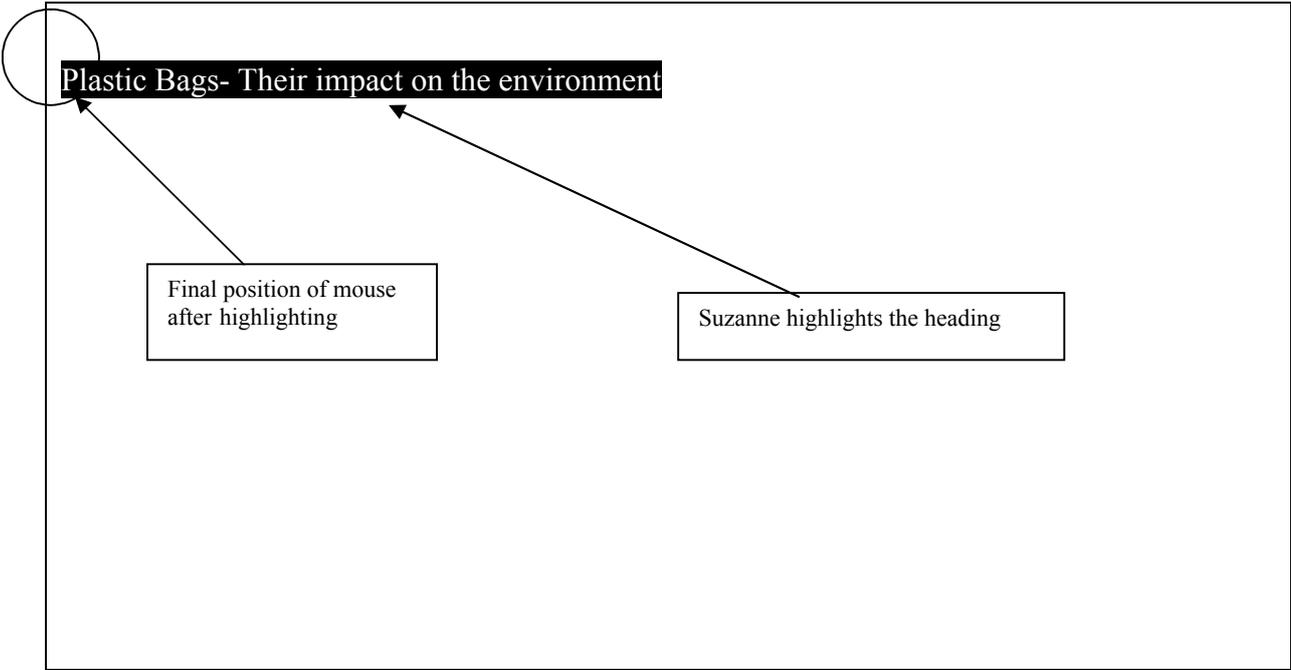
Highlighting the heading

4.1 Discussion

The three extracts discussed above all involve interactional coordination across modalities, specifically between the two participants through talk and through the keyboard/mouse to the computer screen, where they are creating a web page as a classroom assignment. Most of the literature on cross-modal coordination relates to talk and gesture. We find the evidence for the claims we are making from the audio recordings of the two students' voices, the audible tapping of keyboard keys, and the visual tracking of the mouse cursor and vertical flashing cursor on the screen.

Coordination between keyboard/mouse and talk is displayed in a variety of ways and situations.

- **Coordination of speaking aloud the words whilst typing the same words (extract 1)**: We find that speaking is slowed down, in order to map the talk onto the typing; participants start to speak most words slightly before typing them. There are some disruptions of the smooth flow of talk and typing, mostly due to typing errors that are corrected either immediately or a short time later (sometimes the correction is done later during a review of the whole sentence after it has been fully typed). Mostly it is the typist who reads out the word, but we have one instance where the non-typist speaks out the word, which is the last item. The non-typist, like the typist, starts to say the word shortly before she begins typing it (the non-typist, Suzanne, has had a chance by this point to pick up the rhythm and project where the talk is going in order to be able to collaboratively complete the activity).

  In a second example of speaking aloud a word whilst typing **(extract 3)**, it is only the first word that is spoken aloud; as in extract 1, the typing lags very slightly behind the talk.

- **Coordination of reading aloud from the screen and mouse movement (extract 3):** In extract 3, there is coordination between reading a sentence aloud and highlighting that sentence. The action of moving the mouse to the end of the sentence, then sweeping it back across the sentence to highlight it, is coordinated precisely with the reading aloud of the sentence. In this case the reading is, of course,

from left to right, but the movement of the mouse is first from top left to bottom right (coordinated exactly with 'Their impact') and then sweeping and highlighting from right to left (coordinated exactly with 'on the environment').

- **Coordination of talk and bolding using the mouse (extract 2)**: A participant speaks an incomplete sentence, and then suspends it whilst she attends to a different action: She says the incomplete sentence in full coordination, aligning start and finish, with the highlighting of a sentence (which she bolds as her subsequent action), even though what she is saying appears not to be about highlighting and bolding, so there is coordination of talk and action despite an apparent mismatch between the content of the talk and the action. After highlighting and speaking the incomplete utterance, she completes the bolding action in silence. Note also that when she completes the action, the mouse moves around seemingly randomly for 2 seconds, before she begins a next action, namely coordinating talk and changing font size.

- **Coordination of talk and highlighting (extract 2)**: In the first case, this occurs prior to changing the font size using the mouse. Just as the participant begins to utter, 'and then I'll make it a bit bigger', she stops moving the mouse randomly, and sweeps it down to the bottom right of the text to be made bigger, and then sweeps it back over the text (four dot points) to highlight it, preparatory to changing the font size from the toolbar pull-down menu. Again, the action takes longer than the talk, so the remainder of the activity of increasing font size is completed in silence. Once again, when the action is completed, for a short while the mouse moves around rather erratically before the next action and its accompanying talk begin.

- **Coordination of talk and changing font using the mouse (extracts 2 and 3)**: Whilst changing the four dot points from the 'default' font to Chiller font, the participant starts to say, 'I was thinking you know we should change the font for this page', and she starts to type the word Chiller into the font selection window. Just as she finishes this utterance, she strikes the 'Enter' (or 'Return') key on the keyboard, even slightly stretching a sound to achieve this coordination.

On a second occasion, there is coordination of talk with changing the font size; one participant completes the action of increasing the font size of some text (the first heading on the page) from normal to 24pt, whilst the other participant is proposing a next action, namely choosing a different background for the page. Despite this mismatch of action and words, which is by two different speakers, Suzanne completes the action of increasing the font size almost simultaneously with Anna's completion of the utterance.

- **Possible coordination of on-screen activity of correcting typing and talk by the other participant (extract 3)**: Non-typing participant says, 'I don't think' and '(I dunno)' in coordination with the correcting activity.

- **Coordination between the end of typing a sentence by one participant with some talk by the other participant (extract 3)**: The next speaker begins talking just before the prior typist finishes a sentence. This is a case of a speaker self-selecting to take a turn in coordination with what would be a transition relevance place, were the action of the prior participant a turn at talk rather than the typing of a sentence: multi-modal turn-taking, with terminal overlap within the transition space.

5.1 Conclusion

We have seen in this paper how speech and action are coordinated by two year 10 school students in their task of constructing a web page on the computer. The students converse with each other as they engage in discrete activities of adding content step by step to the page, using keyboard and mouse. The students display synchrony and coordination of talk with the emergence of the multimodal text on the screen. As actions on the computer tend to take longer than talking about them (speaking what is being typed, or explaining what one is doing, for example), most coordination of talk and action is at the beginning of the activity; on some occasions, talk and action are coordinated both at the beginning and the end. Second, talk that is coordinated temporally to some action on the screen may be about the action, or may be about something else, but whatever it is, the temporal coordination is precise; in other words, even when action and talk are mismatched in content (the typist is not talking about

what she is doing), talk and action start and finish together. Third, very often the coordination of speaking and the activity on the computer is by one participant acting and speaking unilaterally, but making her talk and action available to her coparticipant. On some occasions one speaker is doing some activity, and it is the *other* participant who coordinates her talk with the actions of the first speaker. Fourth, speaking aloud what is being typed in coordination may not be very surprising, as talk and action are by one speaker with a mapping of spoken words onto the same words in written mode. However, most of the cases of coordination we have are of talk aligned with keyboard or mouse actions without such tight alignment of spoken and written words: That is, this is cross-modal in manner with a talk-technology coordination.

The types of coordination displayed by these students can be seen in hierarchical terms. At a more basic and simple level, there is coordination within one individual between her spoken words and the words as they appear on the screen, with a word-to-word mapping. Second, one individual can provide an account of what she is typing or doing on the screen, where words spoken are not the words being typed. Third, the two students can be talking collaboratively about what one of them is typing or doing on the screen: Thus the non-typing student moves from a role of addressee or recipient to active speaker and producer of talk. Fourth, a student can be typing or doing something on the screen whilst talking about something else, such as planning a later activity that needs to be done, with a mismatch of content of words spoken and action on screen, but nevertheless a temporal coordination of talk and text.

Whilst this coordination of talk and hand/fingers is remarkable, it concurs with previous work on the coordination of talk and gesture. In addition, recent neurological work points to language and action coming from overlapping parts of the brain: Speech and hand actions are processed in the same brain regions. One can draw some parallels in the behaviours of the two students above between 'traditional' writing and working on a computer: The pen plays a similar role to the keyboard and mouse; both are manipulated (in the true sense of the word) by the hands, and they can be seen as equivalent to the vocal tract as producers of language. The paper written upon has parallels to the computer screen, or the waves of sound of spoken language. Further, in computer work, the keyboard is the artefact used for writing, whereas the mouse, with its multidirectional sweeps, is more like gesture. There are, though, some tantalising suggestions that the mouse may be the carrier of new skills. Finally, the kind of coordination we have found in this study is based on two students: There is clearly scope for

27

more research to study in wider and more varied contexts the kinds of coordination we have found.

References

Alac, Morana, 2005. From trash to treasure: learning about brain images through multimodality.. Semiotica 156 (1/4), 177-202.

Alibali, Martha, Kita, Sotaro, Young, Amanda. J, 2000. Gesture and the process of speech production: we think, therefore we gesture. Language and Cognitive Processes 15 (6), 593-613.

Birdwhistell, Ray. L., 1970. Kinesics and Context: Essays on Body Motion Communication. University of Pennsylvania Press, Philadelphia.

Burgoon, Judee, 1994. Nonverbal signals. In: Knapp, M. & Miller, G. R. (Eds.), Handbook of Interpersonal Communication. Sage publications, Thousand Oaks, CA, pp. 229-285.

Capirci, Olga, Contaldo, Annarita, Caselli, Maria Cristina, Volterra, Virginia, 2005. From action to language through gesture: a longitudinal perspective. Gesture 5 (1/2), 155-177.

Chamberlin Quinlisk, Carla, 2008. Nonverbal communication, gesture, and second language classrooms: a review. In: McCafferty, S. G. & Stam, G. (Eds.), Gesture: Second Language Acquisition and Classroom Research. Routledge, London, pp. 25-44.

Choi, Soojung, Lantolf, James P, 2008. Representation and embodiment of meaning in L2 communication: motion events in the speech and gesture of advanced L2 Korean and L2 English speakers. Studies in Second Language Acquisition 30 (2), 191-224.

De Ruiter, Jan-Peter, 2000. The production of gesture and speech. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, New York, pp. 284-311.

De Ruiter, Jan-Peter, 2007. Postcards from the mind: the relationship between speech, imagistic gesture, and thought. Gesture 7 (1), 21-38.

Duncan, Starkey, 1972. Some signals and rules for taking speaking turns in conversation. Journal of Personality and Social Psychology 23 (2), 283-292.

Furuyama, Nobuhiro, 2002. Prolegomena of a theory of between-person coordination of speech and gesture. International Journal of Human-Computer Studies 57 (4), 347–374.

Glenberg, Arthur M., Kaschak, Michael P., 2002. Grounding language in action. Psychonomic Bulletin and Review 9 (3), 558-565.

Goodwin, Charles, 1981. Conversational Organization. Academic Press, New York.

Goodwin, Charles, 2000. Action and embodiment within situated human interaction. Journal of Pragmatics 32, 1489-1522.

Gosling, John, 1981. Kinesics in discourse. In: Coulthard, M. & Montgomery, M. (Eds.), Studies in Discourse. Routledge, London, pp. 148-165.

Gullberg, Marianne, 2006. Handling discourse: gestures, reference tracking, and communication strategies in early L2. Language Learning 56 (1), 155-196.

Hindmarsh, Jon, Fraser, Mike, Heath, Christian, Benford, Steve, Greenhalgh, Chris, 1998. Fragmented interaction: establishing mutual orientation in virtual environments. Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, CSCW, Seattle, Washington, pp. 217-226.

Hostetter, Amos Barr, Alibali, Martha W., Kita, Sotaro, 2007. I can see it in my hands' eye: representational gestures reflect conceptual demands. Language and Cognitive Processes 22 (3), 313-336.

Jefferson, Gail, 1989. Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In: Roger, D. & Bull, P. (Eds.), Conversation: An Interdisciplinary Perspective. Multilingual Matters, Clevedon, pp. 166-196.

Kellerman, Susan, 1992. "I see what you mean": The role of kinesic behaviour in listening and implications for foreign and second language learning. Applied Linguistics 13 (3), 239-257.

Kelly, Spencer D., Barr, Dale J., Church, R. Breckinridge, Lynch, Katheryn, 1999. Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. Journal of Memory and Language 40, 577-592.

Kelly, Spencer, D., Manning, Sarah M., Rodak, Sabrina. 2008. Gesture gives a hand to language and learning: perspectives from cognitive neuroscience, developmental psychology and education. Language and Linguistics Compass 2 (4), 569-588.

Kendon, Adam, 1972. Some relationships between body motion and speech: an analysis of an example. In: Siegman, A. W. &. Pope, B. (Eds.), Studies in Dyadic Communication. Pergamon, New York, pp. 177-210.

Kendon, Adam, 1994. Do gestures communicate? A review. Research on Language and Social Interaction 27 (3), 175-200.

Kendon, Adam, 2004. Gesture: Visible Action as Utterance. Cambridge University Press, New York.

Kita, Sotaro, 2000. How representational gestures help speaking. In McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, New York, pp. 162-185.

Kita, Sotaro, & Özyürek, Asli, 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. Journal of Memory and Language 48, 16-32.

Kopp, Stefan, Bergmann, Kirsten, 2007. Towards an architecture for aligned speech and gesture production. In: Pelachaud, C. et al. (Eds.), IVA 2007, LNAI 4722, Springer-Verlag, Berlin/Heidelberg, pp. 389-390.

McCafferty, Steven G, 2006. Gesture and the materialization of second language prosody. International Review of Applied Linguistics in Language Teaching 44 (2), 197-209.

McNeill, David, 1985. So you think gestures are nonverbal? Psychological Review 92 (3), 350-371.

McNeill, David, 1992. Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago.

McNeill, David, 2005. Gesture and Thought. University of Chicago Press, Chicago.

McNeill, David, Duncan, Susan, 2000. Growth points in thinking-for-speaking. In McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, New York, pp. 141-161.

Namy, Laura L., Newcombe, Nora S., 2008. More than just hand waving: review of hearing gestures: how our hands help us think. Journal of Cognition and Development 9 (2), 247-252.

Pulvermüller, Friedemann, 2005. Brain mechanisms linking language and action. Nature reviews. Neuroscience 6, 576-582.

Schegloff, Emanuel A., 1984. On some gestures' relation to talk. In: Atkinson, J. M & Heritage, J. (Eds.), Structures of Social action: Studies in Conversation Analysis. Cambridge University Press, New York, pp. 266-296.

Schönfeldt, Juliane, Golato, Andrea, 2003. Repair in chats: a conversation analytic approach. Research on Language and Social Interaction 36 (3), 241-284.

Stam, Gale, 2006. Thinking for speaking about motion: L1 and L2 speech and gesture. International Review of Applied Linguistics in Language Teaching 44 (2), 145-171.

Uhlirová, Ludmila, 1994. On the role of the PC as a relevant object in face-to-face communication. Journal of Pragmatics 22, 511-527.

Vygotsky, Lev, S., 1987. Thought and Language. MIT Press, Cambridge.

Willems, Roel M., Hagoort, Peter, 2007. Neural evidence for the interplay between language, gesture, and action: a review. Brain and Language 101 (3), 278-289.

Transcription Symbols

Keyboard

Keyboard |M o s |t _ p |e o p l e _| |d o | e |nt

Letters correspond to the striking of a key on the keyboard. They are lined up with words being spoken when speaking and typing are simultaneous

Keyboard |s̶ g̶ a̶ b̶ _ e̶ i̶ t̶ s̶ a̶ l̶ P |

Strikethrough is used when letters are being deleted on the screen.

↵	This arrow is used to indicate a selection using the 'enter' or 'return' key.
↕	This double headed arrow indicates movement of the flashing vertical bar cursor by using the arrow keys on the keyboard

Mouse

⌐	The 'mouse' symbol is used to indicated the beginning and end of actions carried out using the mouse
⌐	This arrow is used for clicks with the mouse
↗,←, ↖, ↘ ↑,↗,→,↓,↙	These arrows are used to indicate the direction in which the mouse is being moved across the screen, as indicated by the yellow circle cursor
↗↗↗↗↗↗	A series of arrows indicates sustained movement of the mouse in a particular direction (here towards top right across the screen)
BOLD	Indicates selection of 'bold' from the toolbar
DE-HIGHLIGHT	Indicates click on screen to de-highlight text
PDM	Indicates selection of a pull-down menu
FONT SIZE 14	Indicates selection of a font size (here size 14)
PDM-FS	Indicates selection of the 'font size' pull-down menu
PDM-FONT	Indicates selection of the 'font' pull-down menu