International Journal of Lexicography, Vol. 24 No. 1, 2011, pp. 124–153
doi:10.1093/ijl/ecq043 Advance access publication 9 February 2011 

# APPLICATION OF EYE-TRACKING IN EFL LEARNERS' DICTIONARY LOOK-UP PROCESS RESEARCH

Yukio Tono: *Tokyo University of Foreign Studies, Tokyo, Japan (yukio.tono@gmail.com)*

## Abstract

The present study aims to apply eye-tracking technologies to analyse the process of dictionary look-up by learners of English as a foreign language. An experiment was conducted to examine detailed processes of look-up in the microstructure. Several variables (the availability of supporting devices such as signposts or menus, different types of grammar codes, positions of target definitions) were carefully controlled to see how look-up behaviour would change in both monolingual and bilingual dictionary interfaces. The findings show that look-up processes within a microstructure are very complex, showing interactive effects among positions of target information within the microstructure, functions of supporting devices, and users' proficiency levels. Pedagogical and methodological implications will be discussed.

## 1. Introduction

Despite the increasing number of studies on dictionary use, there is one element missing: a detailed analysis of actual look-up processes. Most previous studies used experimental designs in which dictionary use served as an independent variable and was contrasted with no use or other alternatives. The results were always reported by making reference to 'a group with dictionary use' vs. 'no use'. They seldom examined, however, how each user actually looked up the information in a dictionary. This situation looks quite similar to a series of studies on the effects of formal instruction in EFL/ESL classrooms. As Long (1983) critically commented, most studies claimed the effects of 'formal instruction' but very few studies actually described what was happening in real classrooms.

There are several different methodological techniques for describing dictionary look-up processes (see Tono 2001 for a review). Participant observations, self-accounts, think-aloud protocols, videotaping, and screen recorders are some of them. Since dictionary look-up operations are fundamentally cognitive in nature, and most of the information is through visual perceptions, it would be a significant improvement if we could accurately describe the look-up process by eye movements. Eye movement research has been of great interest in neuroscience and psychiatry, as well as ergonomics, advertising and design (Richardson and Spivey 2004). To my knowledge, there has been no attempt to apply this technology in dictionary reference process research.

The present study aims to investigate L2 learners' dictionary look-up processes by using an eye mark recorder. Eye-tracking methodologies seem particularly promising in this area because gaze can be used as a proxy for a user's attention. While many techniques rely on explicit actions of users (e.g., think-aloud, diary reports), eye tracking can yield much more detailed moment-by-moment observations about how users interact with dictionary information. Thanks to this, eye tracking is particularly useful for developing user models in Web sciences and other areas of information technology.

The use of an eye mark recorder has been quite popular in cognitive psychology and ergonomics. Eye movement data have been analysed for two main purposes: *diagnostic* and *interactive*. In the diagnostic use, eye movement data provide evidence of the user's attention and can be investigated to evaluate the usability of interfaces (Faraday and Sutcliffe 1996). In the interactive use, a system responds to the observed eye movements and can thus be seen as an input modality (Duchowski 2003). For instance, an analysis of eye movements in order to assess the usability of an interface for a simple drawing tool was performed in Goldberg and Kotval (1999). Comparing a 'good' interface with well-organized tool buttons to a 'poor' interface with a randomly organized set of tool buttons, the authors showed that the good interface resulted in shorter scan paths that cover smaller areas. The measure of interest in their study is efficient scanning behaviour, i.e. a short scan path to the target object. The merit of this study is to have introduced a systematic classification of different measures based on (temporal) scan paths rather than on cumulative (spatial) fixation areas. The temporal succession of transitions between different areas of attention is particularly relevant to the investigation of the effect of guiding devices such as menus and signposts.

I have conducted a series of experiments (Tono 1984, 1988, 1991, 1992, 2000, 2001) in order to examine in detail what is happening during look-up processes. Methodologically, this study will supplement my previous findings. Using an eye mark recorder will shed some light on new aspects of user behaviour in this complicated process of dictionary look-up.

## 2. Method

### 2.1 *Aims*

The aim of the present study is to investigate the process of dictionary look-up using an eye tracker. To be specific, the following variables are controlled to see if there is any difference in the look-up performances recorded by an eye mark recorder:

(1) Independent variables:
- (a) Interface: monolingual/bilingual
- (b) Position: entry-initial/entry-final
- (c) Guiding devices: menus/signposts
- (d) Information type: definitions/idioms/grammar patterns

(2) Moderator variables:
- (a) Proficiency levels: high/low
- (b) Look-up success: success/failure

(3) Dependent variables:
- (a) Scan paths
- (b) Cumulative fixation areas

It is inevitable that the use of an eye tracker makes look-up performances different from those in normal settings. The experiment was controlled in such a way that the subjects were asked to look at the PC monitor with a head on a chin-rest and search for the information on the display without moving the head. The dictionary information in the entry was carefully manipulated in terms of the variables specified in (1). As indicated in the moderator variables in (2), the present study examines the differences in look-up behaviour between learners at different proficiency levels. The results of an eye-tracking analysis were also assessed in terms of look-up success or failure. Eye tracking data was recorded mainly in terms of scan paths and cumulative fixation areas. The information in the entries was deliberately controlled in terms of its availability and position within the entry in order to see how eye movements will change according to the information provided.

The following is a list of research questions:

(1) How do users search for word meaning in the entry? How does the position of the definition (entry-initial vs. entry-final) affect the process?
(2) How do guiding devices such as menus or signposts affect the process of look-up?
(3) Is the process different if the users' proficiency levels are different?
(4) Is the process different depending on whether the dictionary entry is monolingual or bilingual?
(5) Is there any difference in the look-up process between look-up success and failure?

## 2.2 *Subjects*

Eight subjects (5 female, 3 male) participated in the study. The subjects were selected mainly from students at Tokyo University of Foreign Studies and had learned English for at least six years at junior and senior high schools. They were classified into HIGH and LOW groups, with four persons each, according to proficiency level scores in TOEIC (HIGH: over 800 = B2 to C1 according to the Common European Framework of Reference; LOW: below 550 = A2 according to the CEFR). All the subjects were individually tested in the university laboratory and asked to fill in consent forms for data contribution for research and user profile questionnaires.

## 2.3 *Apparatus*

The presentation of the sample entries was hosted on a computer with a 17 inch (42.5 cm) monitor (the main monitor). A second computer (the EMR monitor) was used to control the eye tracking system, a NAC Image Technology Eyemark Recorder (EMR-8B, model ST-650). The normal unit was used, with which a subject's head is immobilized by a chinrest in order to obtain a high-precision record of an observer's point-of-regard.

The eye mark recorder is shown in Fig. 1, and the experimental setup in Fig. 2.

The EMR eye tracker uses two cameras directed toward the subject's left and right eye, respectively, to detect their movements by simultaneously measuring the centre of the pupil and the position of the reflection image of the infrared LED on the cornea. A third camera is faced outwards, in the direction of the subject's visual field, including the main monitor. The system has a sampling rate of 60 Hz. Eye movement is measured by infrared corneal reflex method in the eye-tracking unit at the frequency of 30 per second. The subject's head posture was maintained with a chin rest, with the eyes at a distance of 50 cm from the main monitor. A digital video recorder that captured the data from


**Figure 1:** The NAC EMR-HM8B and NL8B eyemark recording units used in the study. This figure appears in colour in the online version of the *International Journal of Lexicography*.

**Figure 2:** A general view of the experimental setup. This figure appears in colour in the online version of the *International Journal of Lexicography*.



the third camera was connected to the computer that processed the eye movements and allowed to synchronize eye-tracking recording and video recording.

The overall analysis was made by using a general-purpose analysis software for eye tracking data called the EMR-dFactory. Using the EMR-dFactory, it is possible to obtain processed data such as fixation point, eye blink, angle of convergence, pupil reaction, and gaze data. When eye movements are relatively steady for a short period (250–300 ms), they are called *fixations* whereas rapid shifts from one area to another are called *saccades* (Jacob 1991). During a saccade, no visual processing takes place. In this experiment, the primary focus was on fixation points and scan paths (temporal processes of fixations and saccades).

### 2.4 *Task*

In the experiment, all the tasks were displayed through a PC monitor. Instead of using a real dictionary, special microstructure entries for the words MAKE and FAST were created based on *LDOCE5* and *MEDO* respectively for experimental purposes. Information in the two entries was then deliberately controlled to see whether the effects of availability or non-availability of certain information in the entries would affect actual look-up performance by comparing a set of tasks with and without the given information. In order to answer the research questions (see (1) - (5) in 2.1.), the availability of dictionary information was controlled. Table 1 summarises the list of features controlled in the experiment:

The two entries (MAKE from *LDOCE5* and FAST from *MEDO*) were modified to produce entries with or without the features. Bilingual entries were also written based on the translation of the two entries. This is a design somewhat

**Table 1:** Features manipulated in the task entries

| | |
|---|---|
| (a) | Features available in monolingual entries only: |
| – | [+signpost] / [–signpost]: whether signposts were available or not |
| – | [+menu] / [–menu]: whether menus were available or not |
| (b) | Features available in bilingual entries only: |
| – | [grammar code]: two types of grammar codes were compared |
| (c) | Features controlled in both monolingual and bilingual entries: |
| – | [initial] / [final]: the target definition is in the initial or the final position of the entry |
| – | [HIGH] / [LOW]: the proficiency level of users |

similar to Tono (1984), but the present study focused on eye movements in the search process within a microstructure of an entry.

### 2.5 *Procedure*

The subjects were first briefed about the experiment. They were told that dictionary entries were going to be presented to them on screen, and that they would be asked to find the meaning of the part highlighted in red in the example shown in the upper right edge of the screen. They were also instructed to watch the demonstration carefully so that they could report as soon as they found the information they were looking for. The subjects were then instructed to place their head on the chin rest. Calibration was performed by instructing the subjects to fixate six points on the edges of the screen. After that, the subjects were shown the demo slides, followed by the main session. At the end of the session, they provided some demographic information and answered a short questionnaire about their general English proficiency scores, experience of dictionary use, overseas experience, among others.

### 2.6 *Data Analysis*

Common eye-tracking measures include pupil dilation, fixation information, and sequence information such as scan paths. For the analysis of the present study, I relied on measures related to gaze fixations with a minimum threshold of 100 ms in areas of interest (AOIs). Here AOIs include the structure of the entry and each sub-element therein (e.g., definition, signpost, example, and grammatical codes). Comparisons were made between different proficiency groups in terms of the above gaze pattern and duration data. In the experiment, the subjects' head movements were minimized by a chin support, which made the tasks more artificial than a head-mounted eye tracker, but it made the recording of the eye movements more accurate. The data was analysed using the EMR-dFactory to produce sequence as well as fixation information.

## 3. Results

In the experiment, a close examination was made into eye movements when the users searched for information in the entry. To this end, eight subjects (Subject A – H) were asked to look at an LCD monitor, where a sample entry was displayed with a sentence cue. The sentence contains a word highlighted in red, about which the subjects were asked to look for the information in the entry on screen. Twenty-four entries with controlled information (cf. Table 1) were presented to each subject and the variables manipulated were counterbalanced across the subjects so that any order or carry-over effects were minimized (see online supplementary material for matching). The success-failure was judged against the reports made by the subjects (i.e. the definition or information they chose) at the end of each search.

In this section, I will investigate, using an eye tracker, how the following features affect users' look-up processes: (a) signposts, (b) menus, (c) monolingual vs. bilingual entries, (d) users' proficiency levels and (e) positions of target information in the entry.

### 3.1 *The effect of signposts*

*3.1.1 Overall tendencies.* How do guiding devices such as signposts help? The two contrastive interfaces ([+SIGNPOST] vs [−SIGNPOST]) were compared for the entry MAKE. The following pairs of cue sentences were prepared:

| | | |
|---|---|---|
| (1) | a. John *made* breakfast while we were waiting. | [+signpost] |
| | b. She *made* coffee for us all. | [−signpost] |
| (2) | I *make* that $1200 after tax. | [+signpost]/[−signpost] |
| (3) | a. The novel would *make* a great film. | [+signpost] |
| | b. Good wine can *make* a meal. | [−signpost] |

In most cases, two versions of cue sentence were prepared for each target definition to avoid carry-over effects, although some cue sentences were identical for both conditions (e.g. (2)) to see if the same sentence would be processed in the same way under different conditions.

The overall success rate of the target definition of MAKE, when the variable [signpost] and the proficiency level ([HIGH/LOW]) were controlled, is shown in Table 2. The results show that both groups did not perform well with signposts. Half of the tasks by the high proficiency group ended in failure, although they were provided with signposts. As will be discussed in 3.1.2, an eye tracker confirmed that only one subject from the [LOW] group actually used signposts. However, the [LOW] group subjects did better (n = 8) without signposts, partly because they spent more time on the search and were thus more careful and partly because we could not avoid carry-over effects for some tasks.

**Table 2:** Success rate of look-up performance (MAKE/[±signpost]/[mono])

| MONOLINGUAL | HIGH | | LOW | |
|---|---|---|---|---|
| | Success | Failure | Success | Failure |
| +SIGNPOST | 6 | 6 | 4 | 8 |
| −SIGNPOST | 5 | 7 | 8 | 4 |

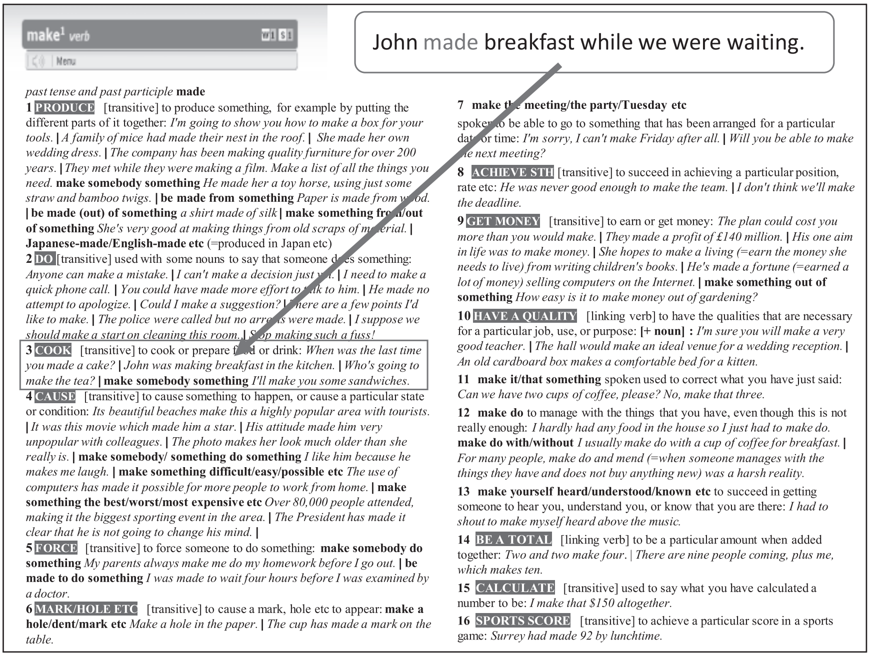
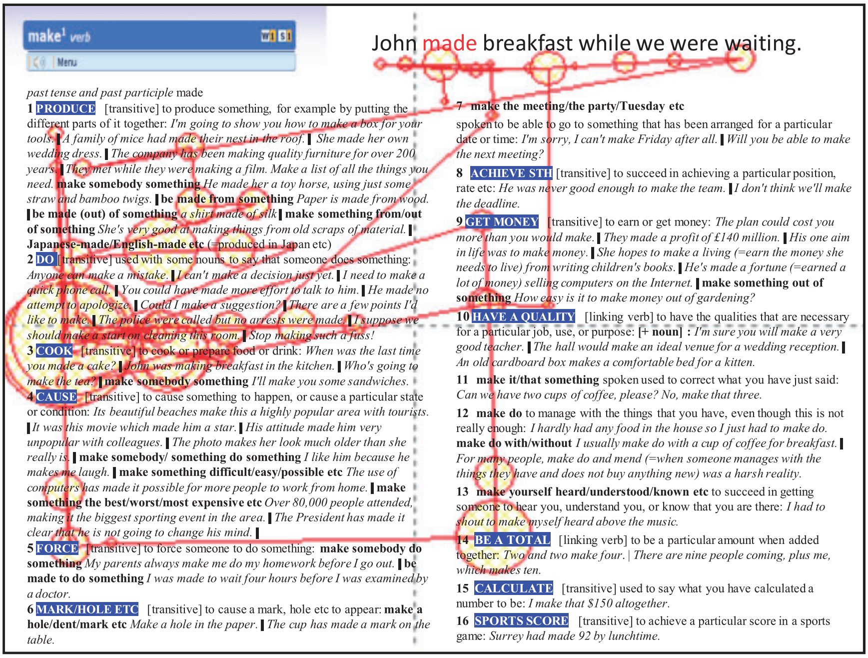



**Figure 3:** Sample entry on screen with a cue sentence and the target definition.

The subjects in the [HIGH] group tended to look up words quickly and sometimes carelessly and often ended up with the wrong definition.

The entry was displayed on screen, as shown in Figure 3, and a cue sentence was given at the upper-right corner with the verb MAKE highlighted in red. The subjects were instructed to look at this cue sentence and search for the target definition which corresponds to the usage in the cue sentence. In this particular example, the subjects were expected to find the third definition as a target. Two versions of the interface with and without signposts were prepared. The scan paths and fixation points were calculated and graphically presented in Figure 4 (A) and (B).

**A**




**B**

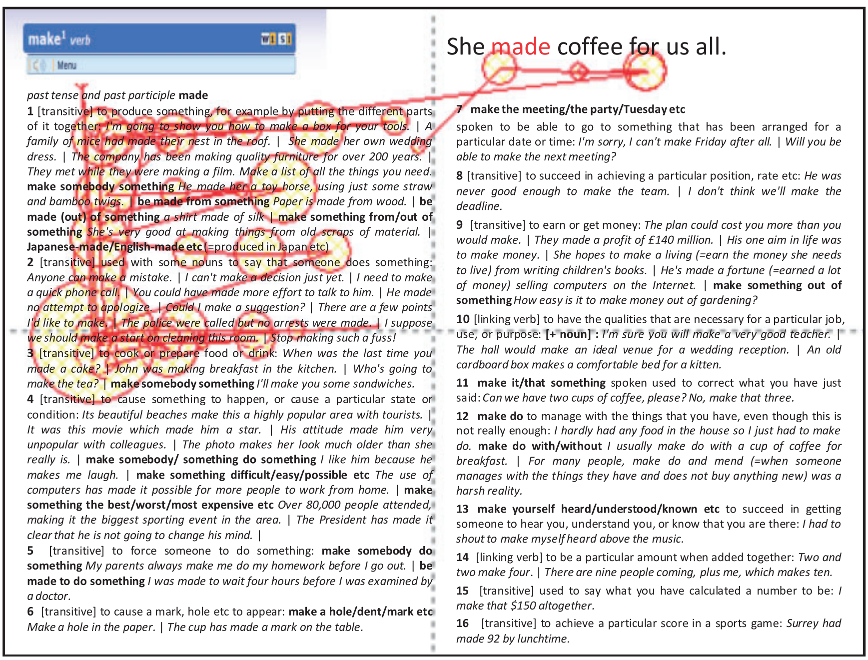



**Figure 4:** Scan paths and fixation points in the entry MAKE (A: [+signpost] B: [−signpost]; see online supplementary material). This figure appears in colour in the online version of the *International Journal of Lexicography*.

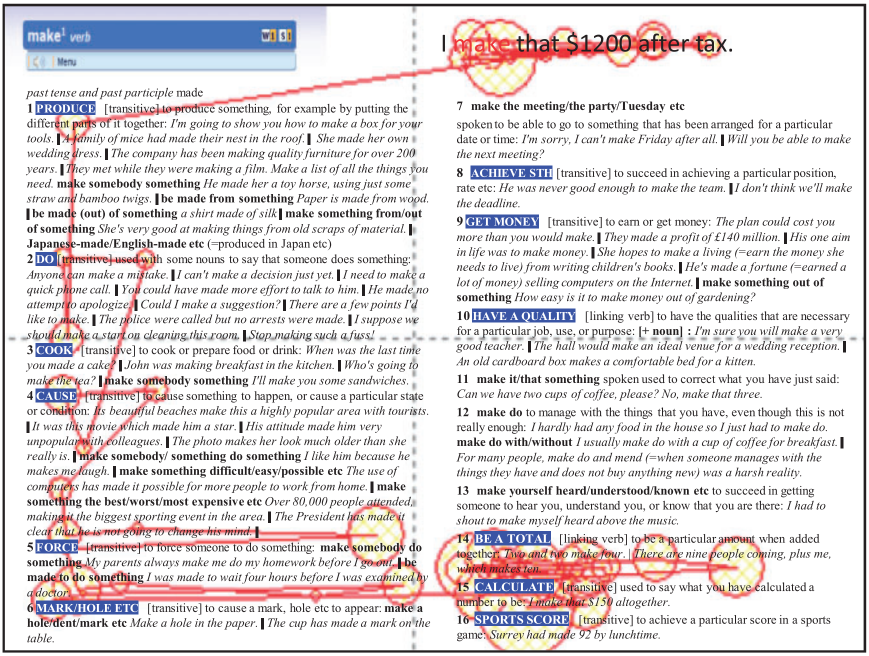
In Figure 4 (A), after looking at the cue sentence, the subject (Subject A: [HIGH]) quickly scanned the signposts in a linear order from the top of the entry and checked the place of the right definition. However, in the case of (B), where no signposts were provided, the same subject started to check each definition by browsing examples one by one to see whether the meanings of MAKE in those examples fit into the given definition. This shows that unless obvious clues such as signposts are provided, the user tends to search the information from the beginning in a rather conservative manner. Finally, in both cases, the subject found the relevant information in Definition No.3, which is the right answer, thus the size of fixation point on that area is very large.

*3.1.2 Signposts and the position of the target definition.* Figure 5 (A) shows the case of an entry with signposts, in which the subjects had to locate the definition toward the end of the entry. The cue sentence is *I make that $1200 after tax*. The user questionnaire indicated that no subject was familiar with this particular usage, and thus this task was different from the previous one described in 3.1. in terms of users' familiarity with the usage. Subject A ([HIGH]) first processed the meaning in the cue sentence and then followed the signposts in a linear order from the top of the entry. Interestingly, he did not move the eyes to the top of the right column, but seemed to find the right signpost (in this case, [CALCULATE]) as he read the last few lines of the left column. Here again, signposts seem to help the users quickly scan the entry structure to get to the information they need.

Figure 5 (B) shows that the same entry was accessed by Subject G ([LOW]), who did not know how to use signposts properly and browsed through the entry from the top and suddenly found the signpost [GET MONEY], read the examples under the definition, and decided on this as a target definition, which turns out to be a wrong choice. Sometimes, the signposts closely related to each other (e.g. [GET MONEY] and [CALCULATE]) may mislead users to a wrong definition. This is especially true if the entry is very long and needs intensive searching. Users want to stop the search as soon as possible, which sometimes ends up with a wrong choice, as shown in Figure 5 (B).

*3.1.3 Signposts and users' proficiency levels.* It should be noted that the effect of signposts was only clearly observed for those who knew how to use them. 3 out of 4 subjects in the [HIGH] group used signposts, while only one subject from the [LOW] group used it. Figure 6 shows the case of Subject E, one of the [LOW] group subjects. She did not use signposts so much and relied mainly on the definitions. It did not take much time to decide on the correct definition, but this is a good example where no reference was made to signposts. The

**A**




**B**

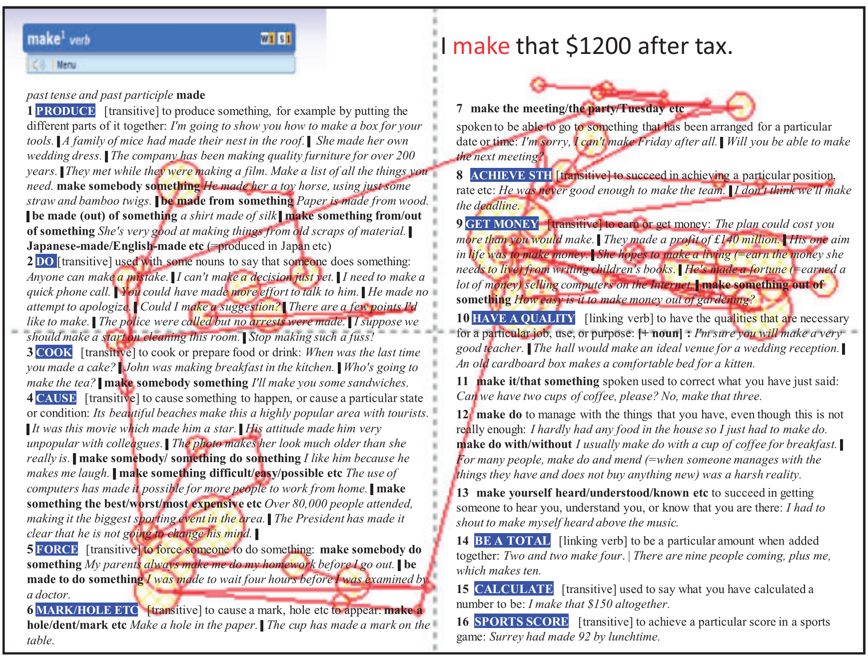



**Figure 5:** Scan paths and fixation points in the entry MAKE (A: [HIGH] B: [LOW]; unfamiliar meanings; see online supplementary material). This figure appears in colour in the online version of the *International Journal of Lexicography*.

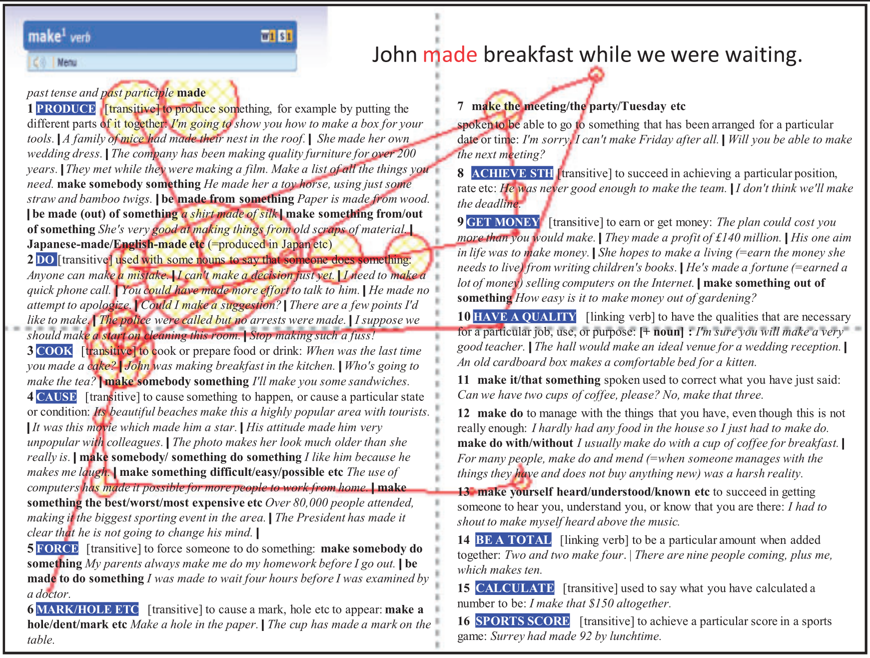


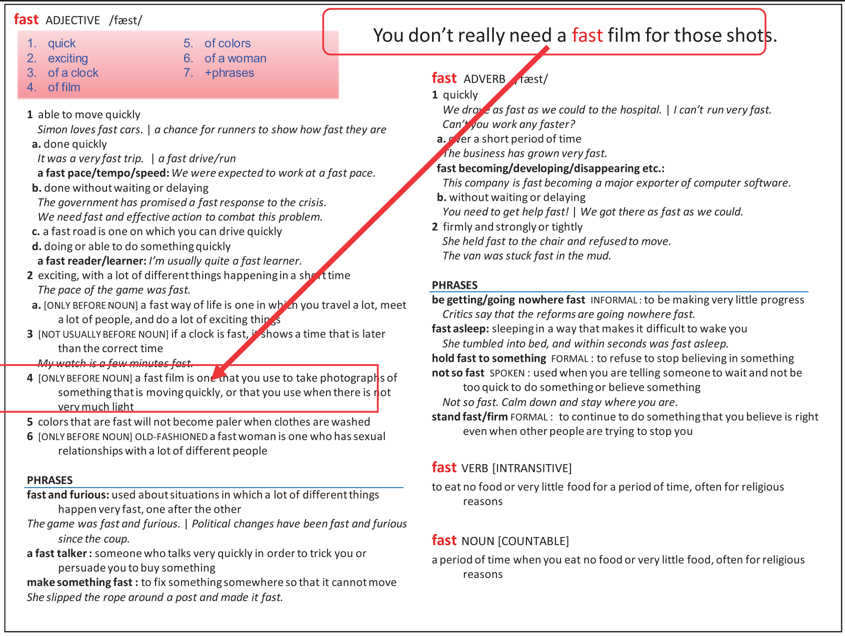

**Figure 6:** The case of no reference to signposts by [LOW] users (see online supplementary material). This figure appears in colour in the online version of the *International Journal of Lexicography*.

[LOW] students did not take advantage of the signposts as much as the [HIGH] students, which suggests that the function of signposts may not be sufficiently transparent to lower proficiency level users, thus deliberate teaching should be needed.

There were also cases in which the subjects in the [HIGH] group, although they used signposts more often than the [LOW] group, did not find a target definition properly. This happened when they relied on signposts too much and did not check against actual definitions or examples. One of the subjects chose the first definition with the signpost [PRODUCE] for *make coffee*, which on first sight is a plausible choice, and she did not choose the target definition with a signpost [COOK]. For her, *make coffee* sounds more naturally related to [PRODUCE] than [COOK]. As was also the case for the subject who selected [GET MONEY] above, signposts sometimes gave too simplistic a view of meaning candidates. Therefore, it is always important that signposts should serve as a 'guide' and need further checking after finding signposted information.

## 3.2 *The effect of menus*

*3.2.1 Effective use of menus.* The menu interface was adopted and modified based on the entry *fast* in MEDO. Figure 7 shows a sample entry with a menu.



You don't really need a fast film for those shots.

**fast** ADJECTIVE /fæst/

1. quick
2. exciting
3. of a clock
4. of film
5. of colors
6. of a woman
7. +phrases

**1** able to move quickly
*Simon loves fast cars.* | *a chance for runners to show how fast they are*
**a.** done quickly
*It was a very fast trip.* | *a fast drive/run*
**a fast pace/tempo/speed:** *We were expected to work at a fast pace.*
**b.** done without waiting or delaying
*The government has promised a fast response to the crisis.*
*We need fast and effective action to combat this problem.*
**c.** a fast road is one on which you can drive quickly
**d.** doing or able to do something quickly
**a fast reader/learner:** *I'm usually quite a fast learner.*
**2** exciting, with a lot of different things happening in a short time
*The pace of the game was fast.*
**a.** [ONLY BEFORE NOUN] a fast way of life is one in which you travel a lot, meet a lot of people, and do a lot of exciting things
**3** [NOT USUALLY BEFORE NOUN] if a clock is fast, it shows a time that is later than the correct time
*My watch is a few minutes fast.*
**4** [ONLY BEFORE NOUN] a fast film is one that you use to take photographs of something that is moving quickly, or that you use when there is not very much light
**5** colors that are fast will not become paler when clothes are washed
**6** [ONLY BEFORE NOUN] OLD-FASHIONED a fast woman is one who has sexual relationships with a lot of different people

PHRASES
**fast and furious:** used about situations in which a lot of different things happen very fast, one after the other
*The game was fast and furious.* | *Political changes have been fast and furious since the coup.*
**a fast talker:** someone who talks very quickly in order to trick you or persuade you to buy something
**make something fast:** to fix something somewhere so that it cannot move
*She slipped the rope around a post and made it fast.*

**fast** ADVERB /fæst/
**1** quickly
*We drove as fast as we could to the hospital.* | *I can't run very fast.*
*Can't you work any faster?*
**a.** over a short period of time
*The business has grown very fast.*
**fast becoming/developing/disappearing etc.:**
*This company is fast becoming a major exporter of computer software.*
**b.** without waiting or delaying
*You need to get help fast!* | *We got there as fast as we could.*
**2** firmly and strongly or tightly
*She held fast to the chair and refused to move.*
*The van was stuck fast in the mud.*

PHRASES
**be getting/going nowhere fast** INFORMAL: to be making very little progress
*Critics say that the reforms are going nowhere fast.*
**fast asleep:** sleeping in a way that makes it difficult to wake you
*She tumbled into bed, and within seconds was fast asleep.*
**hold fast to something** FORMAL: to refuse to stop believing in something
**not so fast** SPOKEN: used when you are telling someone to wait and not be too quick to do something or believe something
*Not so fast. Calm down and stay where you are.*
**stand fast/firm** FORMAL: to continue to do something that you believe is right even when other people are trying to stop you

**fast** VERB [INTRANSITIVE]
to eat no food or very little food for a period of time, often for religious reasons

**fast** NOUN [COUNTABLE]
a period of time when you eat no food or very little food, often for religious reasons

**Figure 7:** Sample entry for FAST with a menu. This figure appears in colour in the online version of the *International Journal of Lexicography*.

There are three pairs of cue sentences for this task:

| | | | |
|---|---|---|---|
| (1) | a. You don't really need a *fast* film for those shots. | [+MENU] |
| | b. You should use a *fast* film in this kind of room. | [–MENU] |
| (2) | a. The fabric was ironed to make the colors *fast*. | [+MENU] |
| | b. It offers a color *fast* guarantee. | [–MENU] |
| (3) | a. The truck was stuck *fast* in the mud. | [+MENU] |
| | b. The boat was stuck *fast* in the mud. | [–MENU] |

Table 3 shows the overall success rate of look-up performance in the entry FAST.

There is a tendency, although it is not statistically significant, that while the [HIGH] group performed fairly well with or without menus, the [LOW] group did much better with menus. This result, however, has to be interpreted with caution. Eye movement analysis revealed that it was only the [LOW] group that actually used the menu. In other words, higher proficiency subjects tended to skip the menu and did not perform much worse for it. In the case of the lower proficiency group, the menu effect was more clearly observed.

Overall, the results seem to suggest that the menu was effective for lower proficiency users, but not advanced learners, which confirmed the results in Tono (1992). Out of four subjects in the [HIGH] group (Subject A–D), only

**Table 3:** Success rate of look-up performance (FAST/[±menu]/[mono])

| MONOLINGUAL | HIGH | | LOW | |
|---|---|---|---|---|
| | Success | Failure | Success | Failure |
| +MENU | 9 | 4 | 9 | 3 |
| −MENU | 8 | 6 | 6 | 6 |

one of them (Subject D) used the menu and the other three bypassed it as they went for a target definition. Figure 8 shows the case in which menus worked efficiently. In Figure 8 (A), after reading the cue sentence, the subject (Subject D; [HIGH]) checked the menu first and successfully located the definition (*adj*. 5 'colors that are fast will not become paler when clothes are washed'). In the case of the three other subjects in the [HIGH] group, however, there was no recorded scan path to the menu section, but they directly moved to the entry. One possible reason may be that some sentence cues contained the use of FAST as an adverb (e.g. *The truck was stuck fast in the mud*.) Since the menu was provided only at the beginning of the adjective use of *fast*, it might be that the subjects in the [HIGH] group first identified the part-of-speech of *fast* in the cue sentence, and moved directly to that section without using the menu (as shown in Figure 8B). In other cases, however, even though the definition was listed in the menu, the subjects in the [HIGH] group did not browse the menu, but started to scan the definition right away. All the four subjects in the [LOW] group, however, used menus whenever provided and the performance was improved with the menus.

Without menus, whether they found the right information largely depends on the complexity of target word usage. For example, the adverb use of *fast*, as in *The truck was stuck fast in the mud*, seems fairly straightforward and very few subjects were lost in the entry. The adjective use of *fast* in the sense of '*colour*', on the other hand, caused a problem, because the usage was unfamiliar to most of them and the definition was relatively short and without an example. Figure 9 shows such a case, where Subject C ([HIGH]) had to read examples in the entry from the top, going back and forth between the cue sentence and the entry. She had difficulty locating the meaning in the entry and it took her almost five times as much time as Subject D with the menu on the same usage.

### 3.3 *Monolingual vs. bilingual entries*

*3.3.1 Overall results.* Monolingual interfaces were also compared with bilingual ones. Bilingual entries were directly translated from monolingual entries. It was hypothesized that the subjects were more used to bilingual entries and could

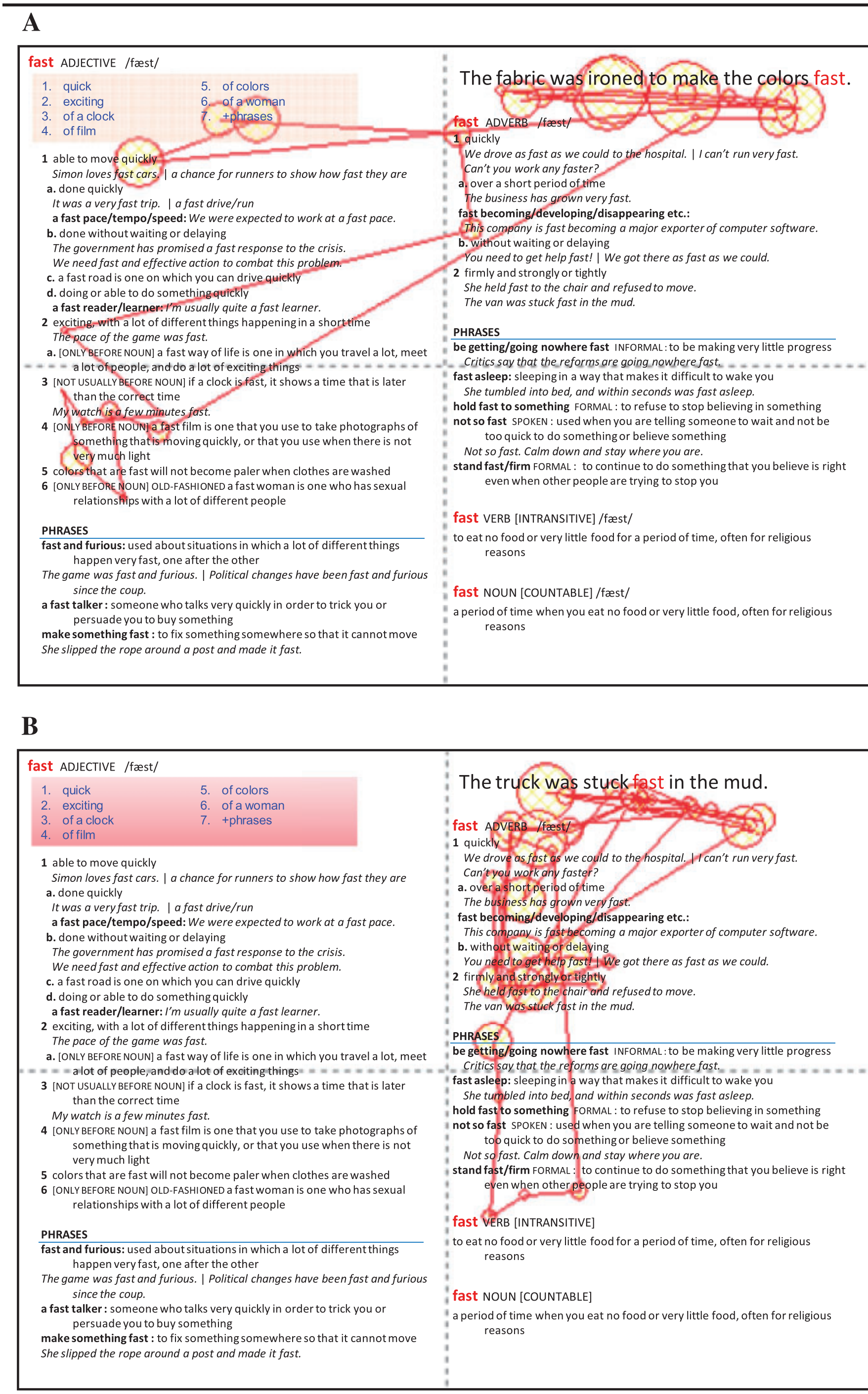



**Figure 8:** Scan paths and fixation points in the entry FAST (A: [+menu] *adj*; B: [+menu] *adv*; see online supplementary material). This figure appears in colour in the online version of the *International Journal of Lexicography*.

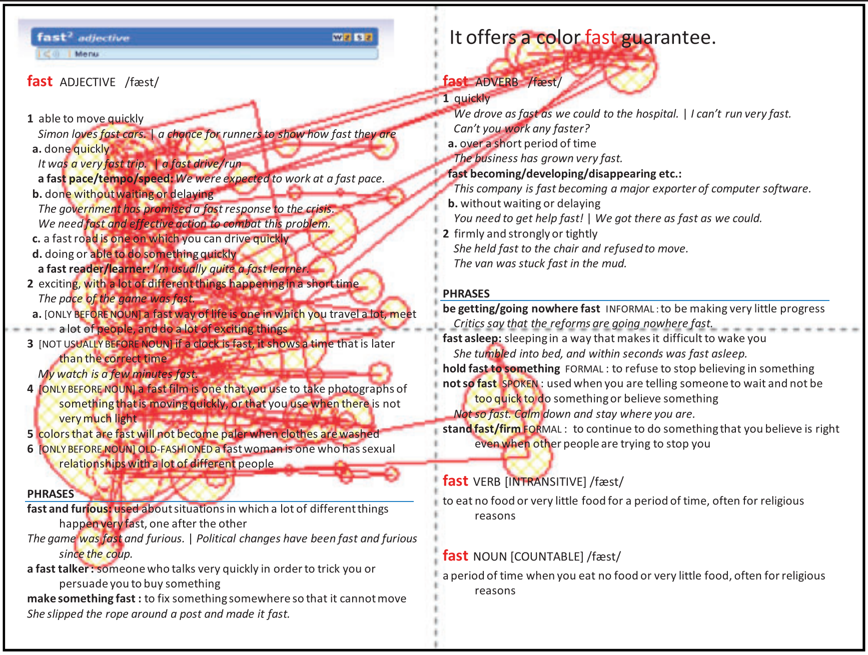



**Figure 9:** Scan paths and fixation points in the entry FAST ([−menu]; see online supplementary material). This figure appears in colour in the online version of the *International Journal of Lexicography*.

retrieve information more quickly. Since translation equivalents, if highlighted in bold, could play a role similar to signposts, bilingual entries were expected to work equally well as monolingual entries with guiding devices. Also, the effect of grammar codes was investigated in comparison to transparent grammar patterns, because grammar codes are still popular in bilingual dictionaries in Japan and it is of interest to examine if users actually use the codes for their search.

Table 4 shows the overall success rates for monolingual vs. bilingual entries. The same entries were compared across different variables, i.e. [±signpost], [±menu] and [mono/bi]. The effect of signposts was not very clear, as was already illustrated in 3.1. Bilingual entries did not seem to help in this respect either. The menu effect, however, was clearly observed in monolingual dictionaries, although menus were mainly used by lower proficiency users. Bilingual entries without menus also had a fairly high success rate.

*3.3.2 Eye-tracker results.* Comparisons between monolingual and bilingual entries in terms of scan paths and fixation paths revealed that it was not bilingual vs. monolingual distinctions that made information retrieval quicker. It depends on the nature of lexical knowledge in question; how salient the information in a cue sentence is, and how straightforward its explanation is in the

**Table 4:** Overall success rate between monolingual and bilingual entries

| | Signpost | | | Menu | | |
|---|---|---|---|---|---|---|
| | Mono | | Bi | Mono | | Bi |
| | +signpost | −signpost | −signpost | +menu | −menu | −menu |
| Success | 10 | 13 | 10 | 18 | 14 | 15 |
| Failure | 14 | 11 | 14 | 7 | 12 | 8 |

entry. Figure 10 shows a pair of tasks in monolingual and bilingual entries. In both cases, the search was fairly simple and quick, because the information sought was rather transparent. Figure 11, however, shows quite a different picture. In both monolingual and bilingual entries, the subject had difficulties locating the target definition properly.

The cue sentence in Figure 11 is ‘The novel would *make* a great film’. This is a linking verb usage of MAKE, which means ‘to be or become something, usually by having the necessary characteristics’ (*CALD*), and this usage is usually listed toward the final position of the entry (in this particular task, sense number 10). It might not be difficult to guess the meaning of the sentence, but it turned out to be quite difficult for users to precisely locate this usage in the entry. This was also the case with the bilingual entry.

*3.3.3 The effect of grammar codes.* The effect of different grammar codes was also investigated. In English-Japanese dictionaries, various grammar codes have been used. For example, the sentence pattern ‘I gave her a present.’ can be expressed by such codes as [SVOO], [make A B], or [+目 + 目] (‘目’ is an abbreviation of ‘目的語 (object noun)’). In monolingual dictionaries, it is now standard practice to avoid codes and make the information as transparent as possible. Thus, most monolingual dictionaries will use the pattern such as [make somebody something] or [make sb sth].

In this experiment, the two coding schemes in bilingual dictionaries were compared: [SVOO] and [make A B]. The former has been adopted in the *GENIUS English-Japanese Dictionary* and others. The latter was first invented by the *ANCHOR English-Japanese Dictionary* and gradually adopted by many learners’ dictionaries in Japan. Whilst the symbols A, B may look odd to native speakers, in Japanese one often explains things by saying ‘A ga B wo nani nani suru’ (i.e. Person ‘A’ does something to Person or Thing ‘B’), and thus prefer to use these initial letters as a coding system. In the case of [SVOO], the code is usually placed right after sense numbers and before translation equivalents. This serves as a flag to locate the structure easily. In the case of the [make A

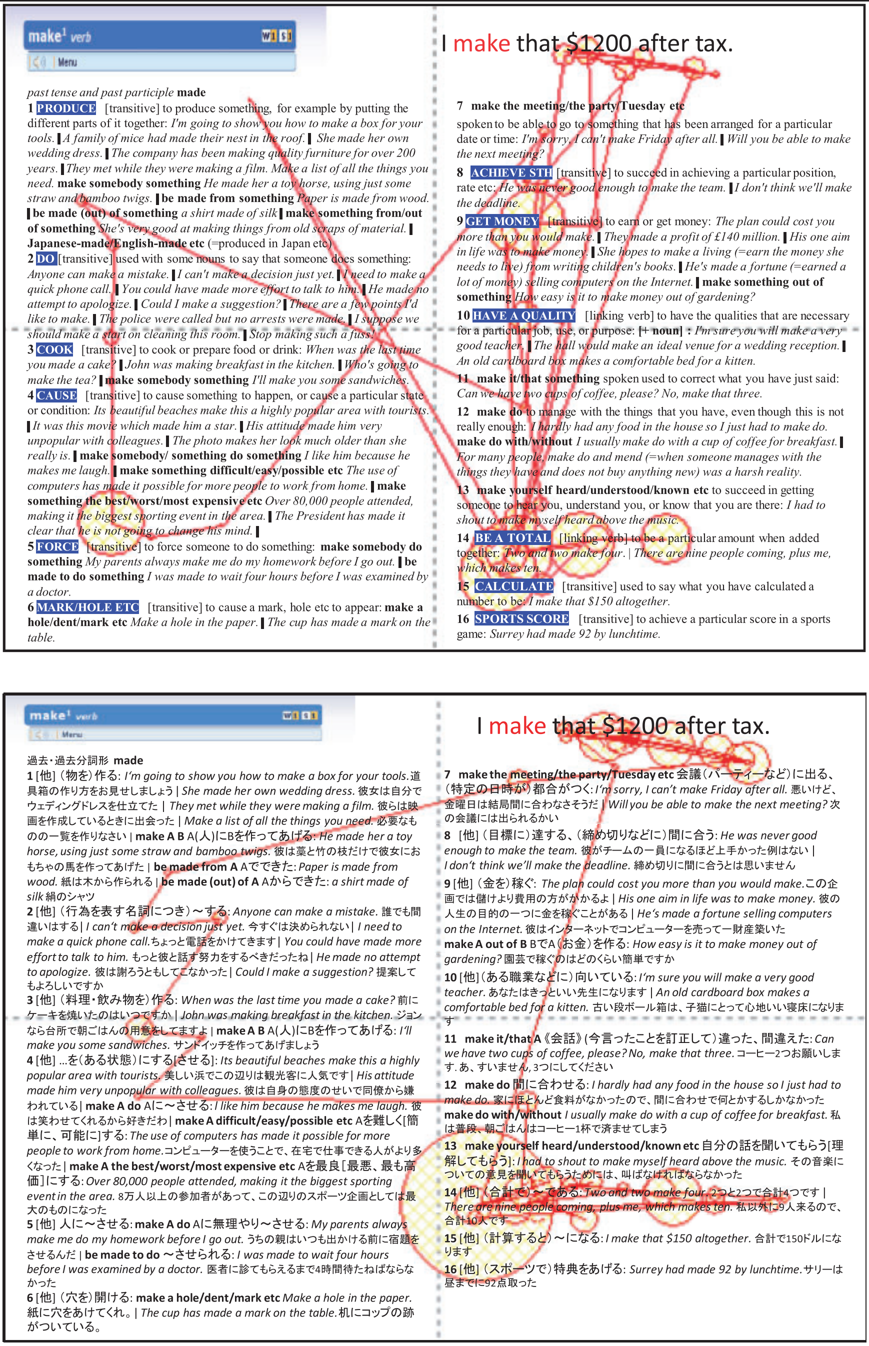



**Figure 10:** Monolingual vs. bilingual entries (Subject C: [HIGH]). This figure appears in colour in the online version of the *International Journal of Lexicography*.

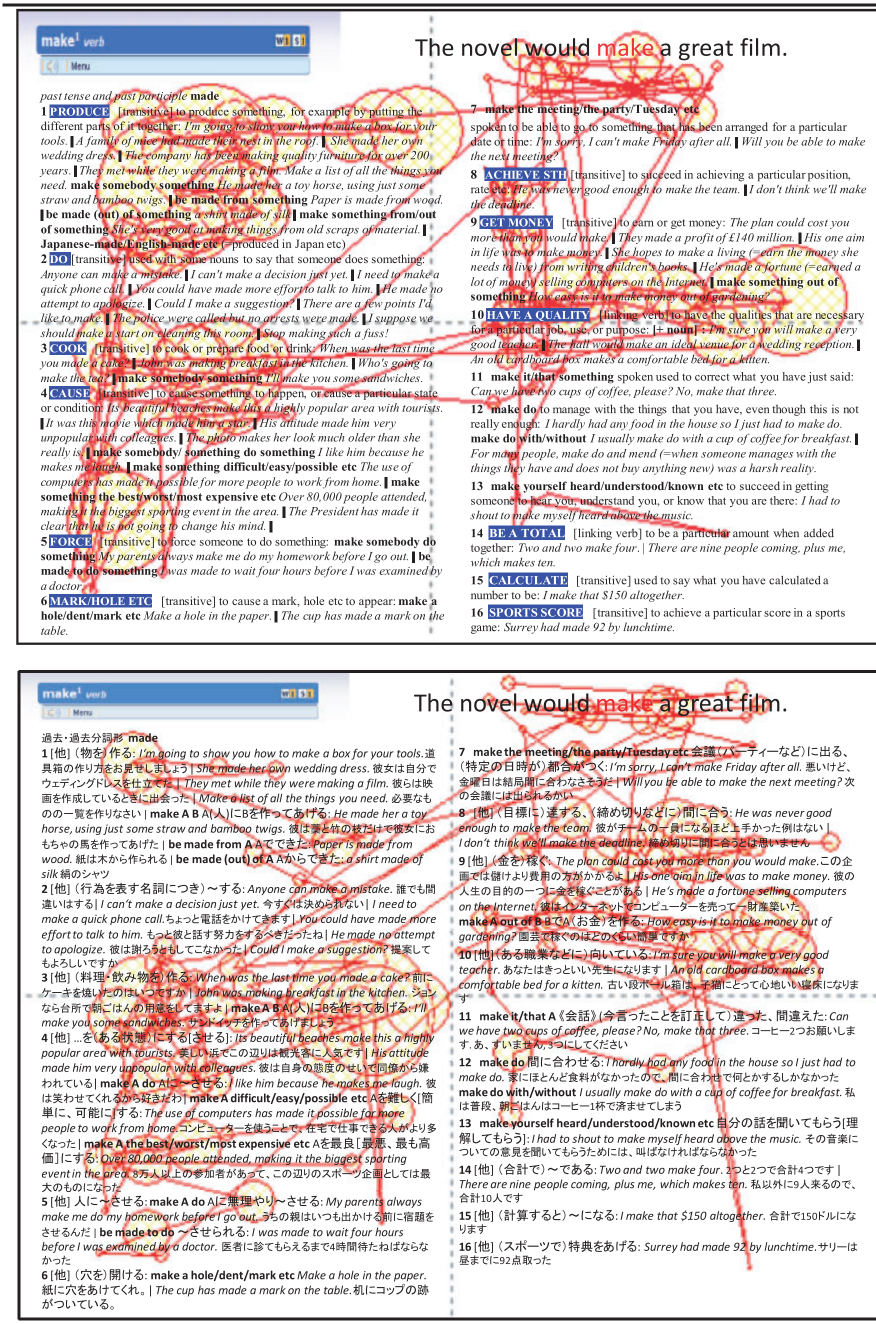



**Figure 11:** Monolingual vs. bilingual entries (Subject N: [LOW]). This figure appears in colour in the online version of the *International Journal of Lexicography*.

B] scheme, it is often placed in front of each structural category with examples. The cue sentences for this comparison are the following:

(1) The prisoners were made to dig holes and fill them again. [SVO do] vs. [be made to do] (Sense No.5)
(2) The haircut made you look ten years younger. [SVOC] vs. [make A do] (Sense No.4)
(3) Butter is made out of milk. [SVO] vs. [be made (out) of A] (Sense No.1)

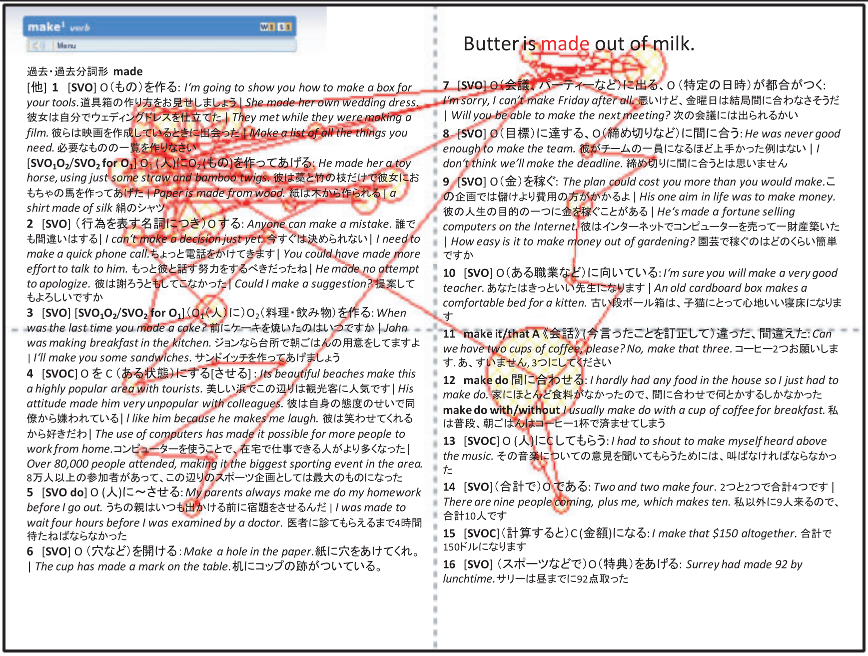
The sentences (1) and (2) are originally the same 'make + obj + bare-infinitive' construction, but *LDOCE5* distinguishes (1) and (2) because of its mandatory, causative sense of MAKE. The two types of grammar codes for each sentence pattern were obtained from the *GENIUS English-Japanese Dictionary* (SVO type) and the *WISDOM English-Japanese Dictionary* (make A type) respectively.

Table 5 shows the results of success/failure in the two grammar codes. The results show that most lookups (close to 80%) were successful. The subjects could manage to locate the target definition correctly. There seems to be no clear difference in the success rate between the two coding schemes. Does this mean that both coding schemes work equally well? The answer is negative. This statistical result is misleading in the sense that it only shows the rate of success in terms of whether the subjects could find the target definition. The analysis of eye mark recorders revealed a strikingly different picture regarding the use of grammar codes.

Figure 12 shows the scan paths and fixation points in the tasks containing grammar codes ([SVOO] type) performed by the [HIGH] and [LOW] subjects (Subject C and E). As shown in Figure 12, it was found that very few subjects used the [SVOO] grammar codes. The results of an eye tracking analysis reveal that only one out of eight subjects accessed the codes and that the rest of them succeeded without using them. Instead of using the codes, they figured out the pattern in a cue sentence and started browsing example sentences. Therefore, even if the success rate was very high, that did not mean that the code itself was useful (see also Dziemianko for more complex aspects of the user-friendliness of grammar codes in this volume).

**Table 5:** Success rates between two grammar coding schemes

| | SVOO type | | make A B type | |
|---|---|---|---|---|
| Proficiency level | Success | Failure | Success | Failure |
| HIGH | 8 | 4 | 11 | 1 |
| LOW | 11 | 1 | 9 | 3 |



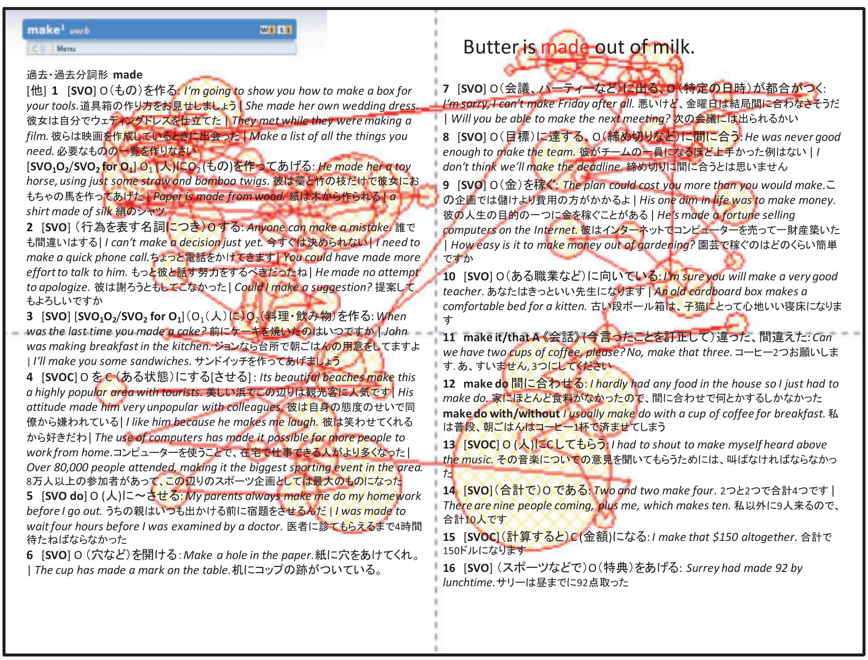



**Figure 12:** Use of the [SVOO] system ([HIGH] vs [LOW] users). This figure appears in colour in the online version of the *International Journal of Lexicography*.

Figure 13 shows the use of the [make A B] type by the two subjects (Subject A: [HIGH]; Subject F: [LOW]). In contrast to the [SVOO] type, the [make A B] type (specifically, [be made (out) of A] in this case) was constantly accessed and guided the subjects to a target sentence. It shows very clearly that this type of grammar coding is more useful for users, compared to the codes at the head of each numbered definition.

To sum up, users behaved differently with the two grammar coding systems. The [SVOO] coding scheme was rarely used in order to locate complex structures in a dictionary. Rather, more transparent codes such as [be made (out) of A] were preferred. Although English monolingual dictionaries do not use the ABC system for grammar codes, the same kind of result would be expected if we compare a scheme such as [be made of sth] against [SVO] or other abstract coding variants. In this sense, transparent codes seem to be the better choice.

### 3.4 *Position in the entry*

Position of target definition in the entry was controlled in order to see the effects of relative positioning in the entry. Tono (1984) found that users only look at the beginning of the entries and tend to be very impatient going through the entry. That is partly the reason why such guiding devices as menus or signposts were invented. Does this tendency still remain the same?

Table 6 summarises the positioning effect of target definition in the entry. Monolingual entries did not show a statistically significant result ($\chi^2 = 0.99$, n.s.), while there was a significant difference in the success rate between the two positions (entry-initial vs. entry final) of the target definition in the case of bilingual entries ($\chi^2 = 11.65$, $p < 0.001$). However, a closer look at the cases of monolingual dictionaries in terms of proficiency levels revealed that the odds ratio for the [HIGH] group's success rate in the entry-initial position was much higher than the [LOW] group (odds ratio = 1.71; Odds ratio is the estimate of $(p_1/(1\text{-}p_1))/(p_2/(1\text{-}p_2))$, where $p_1$ is the number of success by the [HIGH] group and $p_2$ is that by the [LOW] group = 1.71), which suggests that lower proficient users have difficulties retrieving information even in the entries where target definitions were located close to the beginning. In the case of bilingual dictionaries, the success rate for entry-initial items was dramatically improved, which suggests that bilingual entries are easier to process for users.

The final row of Table 6 shows the information on idioms and fixed phrases. In this experiment, the number of observations turned out not to be sufficient, and thus no conclusive remark could be made. However, there is an interesting pattern of scan paths, as shown in Figure 14. When the subject looked at a cue sentence 'Can you make yourself understood in Russian?', she immediately moved down to the end of the entry and select the phrase 'make yourself heard/understood/known'. She somehow knew that this is a fixed expression and that a dictionary will cover this as a phrase. Usually phrases or idioms are

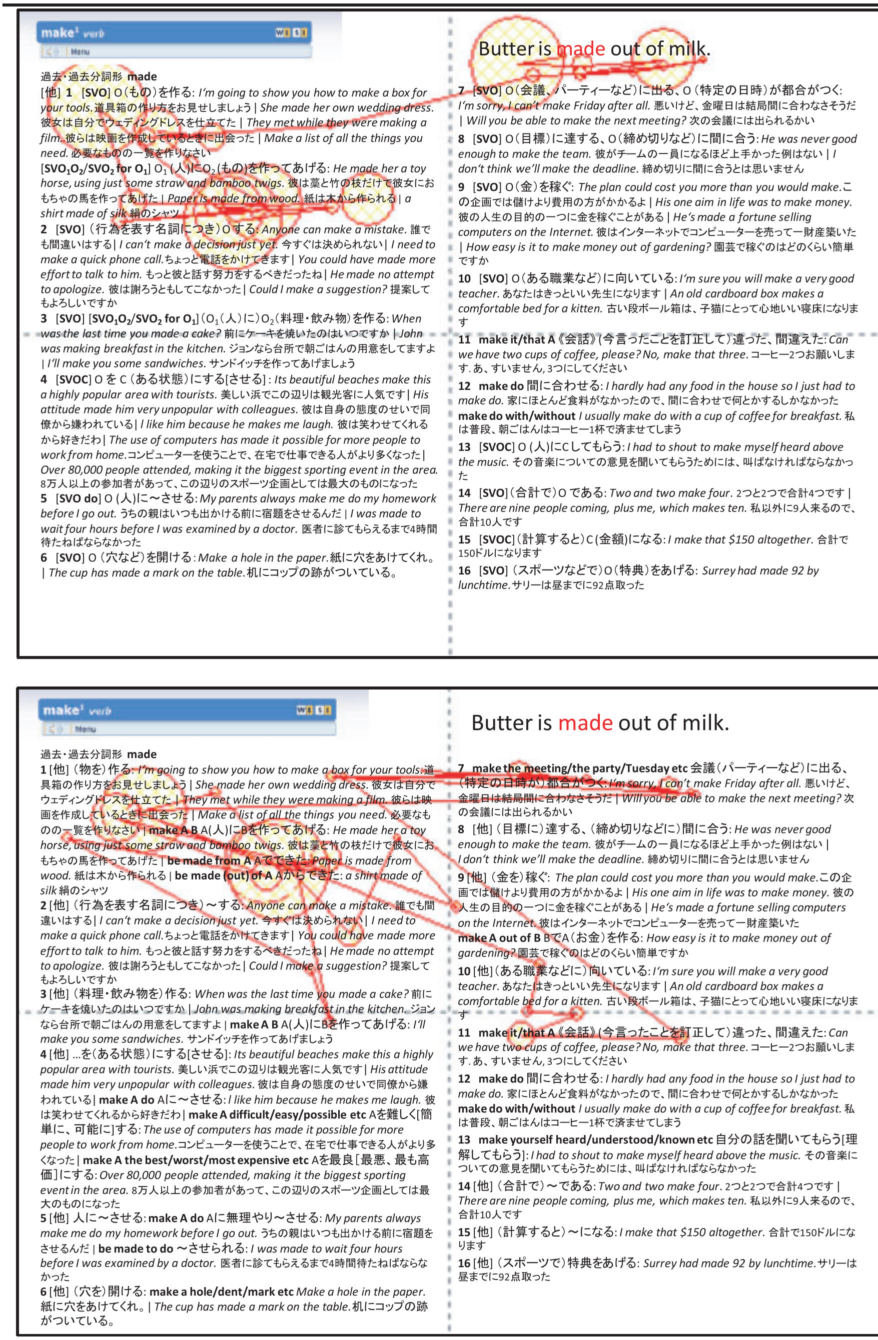



**Figure 13:** Use of the [make A B] system ([HIGH] vs. [LOW]). This figure appears in colour in the online version of the *International Journal of Lexicography*.

**Table 6:** The effect of positions of target definition in the entry

| | Monolingual | | Bilingual | |
|---|---|---|---|---|
| | Success | Failure | Success | Failure |
| Long | 28 | 29 | 28 | 26 |
| Short | 20 | 12 | 35 | 5 |
| Phrase | 7 | 3 | 1 | 0 |

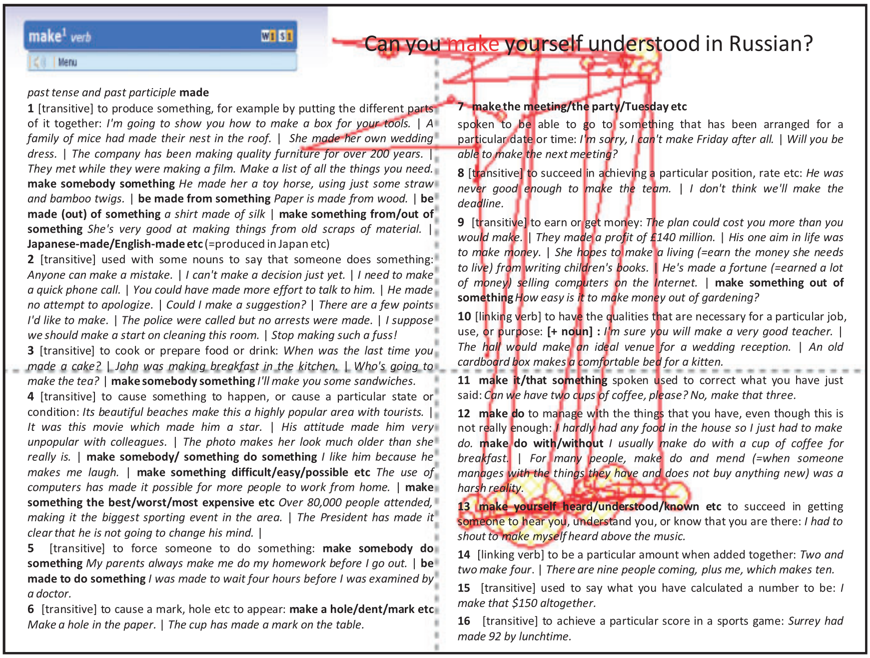



**Figure 14:** The scan paths for phrase/idiom search (Subject G: [LOW]). This figure appears in colour in the online version of the *International Journal of Lexicography*.

treated after main definitions. Therefore, users can jump to phrase/idiom sections as soon as they figure that it should belong there.

### 3.5 *Summary*

Let me summarise the overall results by showing the mosaic plot of three variables ([monolingual/bilingual], [information type], and [success/failure]). See Figure 15 for the results. Vertical lines between the boxes mean that there were no corresponding observations for that variable. The performance regarding signposts (see the areas named 'sp(+)' vs. 'sp(–)') show mixed

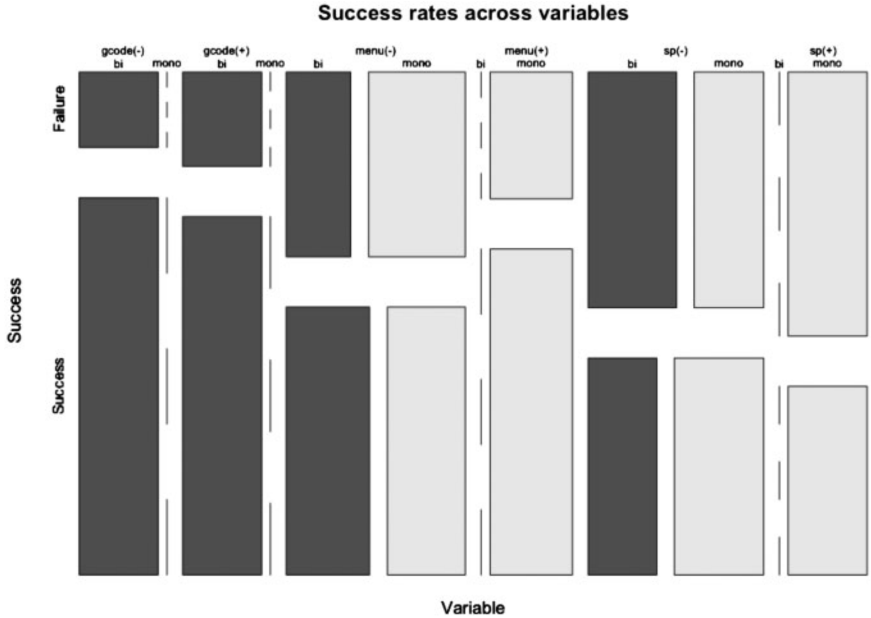



**Figure 15:** Success rates across variables ([information type] x [mono/ bilingual]).

results. The eye tracker confirmed that the [HIGH] group used signposts more than the [LOW] group, but that did not lead to better results. Some users, after browsing signposts, selected a wrong definition as a target, mainly because some signposts did not work as efficiently as expected. If the meanings of signposts are close to each other (e.g. [GET MONEY] and [CALCULATE] in the entry MAKE), they sometimes interfere with the user's right decision. Some signposts were too abstract (e.g. [HAVE A QUALITY]) and it is difficult to infer the meaning from it.

My hypothesis was that bilingual entries would work better than signposts in monolingual dictionaries, but this hypothesis was not supported (see 'bi' under 'sp(−)'): at least an equal rate of failure was observed in bilingual entries as well. This means that some types of lexical information are very difficult to retrieve from a dictionary. The simple solution of providing signposts is not sufficient.

The menu effects, on average, were more salient than signposts. As was illustrated in 4.2, the menu effects were especially significant for the lower proficiency users, which confirms the results in Tono (1992). The higher proficiency group did not make any difference in performance with or without menus.

The two types of grammar codes were tested for their effectiveness. Although the results show relatively high success rates for both types of coding schemes, the picture was rather simplistic. Eye movement analysis shows that most subjects ignored [SVOO] type codes, and did not use the information at all when searching for the definition. The embedded codes ([make A B] type) at the beginning of each example set, however, were found to be constantly accessed and seemed to work fine for both levels of users.

## 4. Discussion

Dictionary reference skills are complex, integrated skills which involve language skills, knowledge about dictionary conventions, problem-solving skills, information processing skills among others. Users have to identify encoding/ decoding problems in context, and make a decision about problem-solving strategies. In other words, they have to decide whether to guess from context, consult a dictionary, or simply ignore it and continue the task. If users decide to consult a dictionary, then the next task is to decide what information to look for in which dictionary. Here users should have a basic understanding of the language and dictionaries. Skilled users will employ various linguistic as well as contextual clues, such as guessing part of speech from the context, judging whether it is part of idiomatic expressions or not, collecting grammatical information such as count or mass nouns, verb complementation patterns, inflected or derived words, etc. Then with certain assumptions, they look up a particular word in question, go to the headword, and search for the information that answers to your question. At every step, users have certain expectations about what to see or get. All the query results will be judged against the criteria of whether it will meet the expectations or not.

Speculating on the process of dictionary look-up in this way might be useful to build up a theoretical model of dictionary reference behaviour, but actual look-up processes are far more complicated and unpredictable than this. In the present study, the use of an eye mark recorder has made it possible for the first time to make a detailed description of the paths that users went through while searching for the information in a dictionary, and what information they paid attention to in terms of points of regard. Here, major findings will be briefly reviewed and further interpretations and implications will be discussed.

The experiment examined in detail the search process in the microstructure of an entry. The eye tracking system revealed a very interesting picture of what was actually happening during the process. First of all, the findings show that about two thirds of attempted searches resulted in success. This means that one third of the searches resulted in failure. The subjects tried to look for information in a dictionary, but more than 30% of them ended up with wrong answers. This happened even in cases where they were given guiding devices in the form of menus or signposts. The scan paths in failed attempts show two typical patterns; it could be a very simple path, where users jumped to particular definitions immediately after seeing the cue, or an extremely complex, tangled path, where users moved their eyes all over the entry but could not find an answer. This finding has two implications. First, even though users consult a dictionary, that does not automatically mean that they bring back the right information with them. We should give hands-on practice to learners as dictionary users about how to look for information properly in a dictionary in order to make sure that their search attempt will result in more success. Second, failure does not simply mean a waste of time. The scan paths of failed attempts indicate that the subjects spent so much time processing the content of the entry to find the answer. Even if the final answer was wrong, reading all the examples and figuring out the meanings may contribute to more learning.

The second major finding of the eye movement analysis is that there are complex interactions between look-up support devices (e.g. signposts or menus) and users' skills and language proficiency levels. The results show that menus were only effective for lower proficiency users. Higher proficiency users simply skipped the menus. This confirms the previous findings in Tono (1992). Lew and Tokarek (2010) found that the menu with highlighting is better, but their dictionary interface is an online version, where pressing the menu will take the users directly to the definition. The menu function would be rather different from the paper version, and their findings should be interpreted carefully. Signposts, on the other hand, were accessed mostly by higher level users. Lower proficiency users did not look at them simply because they did not understand the functions of signposts. Lew and Pajkowska (2007) also examined the effects of signposts on access speed and look-up task success by two groups of different educational levels and found only a significant difference in access speed. There is a possibility that some of the subjects in their study did not use signposts at all. Since some phrases such as 'as black as pitch' were highlighted (shown in the Appendix in Lew and Pajkowska 2007), as in grammar codes in my study, the subjects might have preferred to search for the highlighted target phrase through the main body of the entries. This kind of observation, however, can only be confirmed with a device such as an eye tracker.

Even those higher proficiency users who used signposts sometimes failed to get to the target definition. This happened when signposts were either too abstract or overlapped with other signposts, and thus were confusing. Therefore, it is desirable to empirically test the effectiveness of those devices and improve their quality (see Nesi and Tan, in this volume). The implication is that this kind of supporting device is supposed to help users, but in many cases users are conservative and unwilling to use those unless instructed. Menus are more transparent than signposts in this sense. Deliberate teaching is necessary whenever new devices are introduced. Also some devices are closely connected to particular levels of user's language as well as reference skills. Since these devices take much space, a decision has to be made about whether to introduce them, depending on the skill levels of target users.

Thirdly, a comparison between monolingual and bilingual entries revealed that whilst scanning bilingual entries is basically much easier than monolingual ones, bilingual interfaces did not outperform monolingual ones in searching for complex lexical information. If the information is located at the beginning of an entry, the success rate of retrieving correct definitions is much higher with the bilingual interface. However, if the information is located at the end of an entry or if it is not obvious, bilingual entries did not seem to help much. Previous studies show that users always prefer bilingual dictionaries to monolingual. While the bilingual dictionary has obvious advantage, it is also true from this experiment that bilingual/monolingual distinctions do not matter that much when it comes to searching for complex lexical information.

Fourthly, the results show that it is sometimes misleading to compare groups with or without particular dictionary information. In the present study, a comparison was made between entries with two different grammar codes. Apparently, the results seemed to show that there was no significant difference between the two grammar coding systems. However, the eye tracking analysis clearly shows that one of the coding systems was not accessed at all. I think this is what happened in many studies on dictionary use. While they investigated the effects of dictionary use vs. no use, they did not examine how the subjects used dictionaries, for what purposes and for which words. They performed significance testing between the groups and found a difference, but in many cases they were not certain what aspect of dictionary use contributed to the difference. Without a closer examination of the actual look-up operations, it is difficult to make any generalization from such studies regarding the effects of dictionary use.

Finally, a word of caution is in order here with respect to the nature of this kind of research. While the eye mark recorder is a powerful tool, the setting inevitably becomes artificial. In order to calculate gaze points accurately, it was necessary to fix the subjects' head onto the chinrest and ask them to look at the PC monitor, instead of real dictionaries. This could elicit some unnatural behaviour from the subjects. Also they worked on twenty-four different versions of the two entries, MAKE and FAST. Although the tasks were counter-balanced across subjects, there were bound to be some carry-over effects. The combination of target lexical information and the entries which contain that information needs to be carefully designed. Individual differences are also a serious issue in this type of research. Since the number of subjects in this kind of research has to be limited, individual differences in response patterns are hard to control. Overall, the approach itself is quite promising, but a more solid design and its replications with different subjects or exploitation of a single-subject design would definitely be needed.

## 5. Conclusion

The present study has been the first attempt to employ an eye tracker to precisely catch the eye movements of dictionary users in search of information in a dictionary entry. In spite of its artificial nature, recording detailed scan paths and gaze points has been revealing. It also shows a very complex nature of look-up processes, affected by the content of microstructure, guiding devices, information types, users' language proficiency levels, and skill levels. The results also suggest that more research into actual look-up processes is needed in order to test the claims related to the effectiveness of dictionary use in language learning and teaching. I hope that this study will trigger more interest in taking rigorous research methods using such apparatus as an eye mark recorder.

## Acknowledgements

I thank the editor of the special issue, Robert Lew, for helpful comments and discussions. I also thank the editor of IJL, Paul Bogaards, for his comments on my early draft. Thanks also go to my postgraduate student, Taku Kaneta, for devoting his time to the operation of the eye tracker and the data analysis. The usual disclaimers apply.

## Supplementary material

Supplementary material (in the form of movie files) is available at *International Journal of Lexicography* online.

## References

### A. Dictionaries

**Inoue, N. and I. Akano (eds) 2006.** *The WISDOM English-Japanese Dictionary*. (Second edition). Tokyo: Sanseido. (WISDOM).

**Konishi, T. and K. Minamide (eds) 2001.** *Taishukan's GENIUS English-Japanese Dictionary*. (Third edition). Tokyo: Taishukan. (GENIUS3).

***Macmillan English Dictionary Online*.** (MEDO), 15 January, 2010. http://www.macmillandictionary.com/.

**Mayor, M. (ed.) 2009.** *Longman Dictionary of Contemporary English*. (Fifth edition). London: Longman. (LDOCE5).

**Shibata, T., S. Kobayashi, and T. Konishi (eds) 1972.** *The ANCHOR English-Japanese Dictionary*. Tokyo: Gakushu Kenkyu Sha. (ANCHOR).

### B. Other references

**Duchowski, A. T. 2003.** *Eye Tracking Methodology: Theory and Practice*. London: Springer.

**Dziemianko, A. 2011.** 'User-friendliness of Noun and Verb Coding Systems in Pedagogical Dictionaries of English: A Case of Polish Learners.' In this volume.

**Faraday, P. and A. Sutcliffe. 1996.** 'An Empirical Study of Attending and Comprehending Multimedia Presentations' In *Proceedings of ACM Multimedia 96*. Boston MA, 265–275.

**Goldberg, J. H. and X. P. Kotval. 1999.** 'Computer Interface Evaluation Using Eye Movements: Methods and Constructs.' *International Journal of Industrial Ergonomics*, 24.6: 631–645.

**Jacob, R. J. K. 1991.** 'The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at Is What You Get.' *ACM Transactions on Information Systems*, 9.3: 152–169.

**Lew, R. and J. Pajkowska. 2007.** 'The Effect of Signposts on Access Speed and Lookup Task Success in Long and Short Entries. *Horizontes de Lingüística Aplicada*, 6.2: 235–252.

**Lew, R. and P. Tokarek. 2010.** 'Entry Menus in Bilingual Electronic Dictionaries'. In S. Granger and M. Paquot (eds), *Elexicography in the 21st Century: New Challenges, New Applications*. Louvain-la-Neuve: Cahiers du CENTAL, 193–202.

**Long, M. 1983.** 'Does Second Language Instruction Make a Difference? A Review of Research' *TESOL Quarterly*, 17.3: 359–382.

**NAC. Image Technology, 2004.** URL: http://eyemark.jp.

**Nesi, H. and K. H. Tan. 2011.** 'The Effect of Menus and Signposting on the Speed and Accuracy of Meaning Selection'. In this volume.

**Richardson, D. C. and M. J. Spivey. 2004.** 'Eye-Tracking: Characteristics and Methods. In G. Wnek and G. Bowlin (eds), *Encyclopedia of Biomaterials and Biomedical Engineering*. New York: Informa Healthcare, 568–572.

**Tono, Y. 1984.** *On the Dictionary User's Reference Skills*. Unpublished B. Ed. Thesis. Tokyo: Tokyo Gakugei University.

**Tono, Y. 1988.** 'Can a Dictionary Help One Read Better?' On the Relationship between E.F.L. Learners' Dictionary Reference Skills and Reading Comprehension. In G. James (ed.), *Lexicographers and Their Works*. Exeter: University of Exeter Press, 192–200.

**Tono, Y. 1991.** 'A Good Dictionary User: What Makes the Difference? In K. Ito, K. Kanatani and T. Noda (eds), *Recent Studies on English Language Teaching*. Tokyo: Yumi Press, 229–253.

**Tono, Y. 1992.** 'The Effect of Menus on EFL Learners' Look-up Processes' *LEXIKOS*, 2: 230–253.

**Tono, Y. 2000.** 'On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks. In U. Heid, S. Evert, E. Lehmann and C. Rohrer (eds), *Proceedings of the 9th EURALEX International Congress, EURALEX 2000*. Stuttgart: Institut fur Maschinelle Sprachverarbeitung, Universitat Stuttgart: 855–862.

**Tono, Y. 2001.** *Research on Dictionary Use in the Context of Foreign Language Learning*. Lexicographica Series Maior 106. Tübingen: Max Niemeyer Verlag.