# Case-based Natural Language Dialogue System using Facial Expressions

**Satoko SHIGA**
Fujitsu Laboratories LTD.
1-1, Kamikodanaka 4-chome,
Nakahara-ku, Kawasaki,
211-8588, Japan
shiga.satoko@jp.fujitsu.com

**Seishi OKAMOTO**
Fujitsu Laboratories LTD.
1-1, Kamikodanaka 4-chome,
Nakahara-ku, Kawasaki,
211-8588, Japan
seishi@jp.fujitsu.com

## Abstract

A new method for case-based natural language dialogue system is presented. This system deals with not only the utterance sentences that are used in usual case-based dialogue systems, but also facial expression information to express past cases. As a result, it can improve the appropriateness of response and present the system's utterance along with facial expression information to the user. We show the advantage of our system over other systems by using examples of dialogue provided by our system.

## 1 Introduction

A natural language dialogue is one of the best ways for creating a man-machine interface. Although many approaches for dialogue systems have been proposed, including a template-based approach (Weizenbaum, 1966) and a plan-based approach (Allen et al., 1994; Carberry, 1990), in this paper, we apply a case-based approach.

Case-based reasoning (CBR) is a reasoning model that solves a new problem by using previous observations. Past cases, which consist of pairs of problems and their solutions, are stored in a case-base. The system recalls a similar case to the new problem, and then the solution of the selected case is modified to adjust to any difference between the new problem and the past problem. Finally, the system puts forward the modified solution as the solution to the new problem. CBR has the following advantages over other approaches (e.g.,(Leake, 1996)):

- The cost of knowledge acquisition is low, because the system only has to record facts that actually happen as cases.
- Knowledge maintenance is easy because the system learns incrementally. The cases are added automatically, and it is unnecessary to take into account the consistency of knowledge.
- Quality of solutions is increased even though the domain is ill-defined, because the system can treat phenomena that are difficult to formalize.
- Problem-solving efficiency is increased because the system gets shortcut to the successful solution by reusing the case.

In applying the CBR model to dialogue systems, a past dialogue history is stored as a case in a case-base. To generate a response, the system retrieves a similar utterance to the current context from the case-base, and modifies the response utterance of the case to suit to the current situation.

In making a dialogue system, the advantages of CBR are important for the following reasons:

- A large quantity of complicated templates or planning rules must be used in the template-based or plan-based dialogue system. It is, however, quite difficult to make an enough quantity manually. The case-based approach reduces the cost of the knowledge acquisition

and makes possible the system construction easily.

- Knowledge maintenance is a thorny issue in other approach. To develop the system's vocabulary, for example, it is often necessary to revise the whole rule (because adding one rule often means rewriting a large part of the rules). In case-based approach, we just have to add cases of utterances including the new word.

- There are various ways to respond to one utterance, and it is difficult to formalize them as rules. The case-based approach is suitable for such domain to provide the high-quality solution.

- The template-based or plan-based systems can not deal with unexpected dialogues. In contrast, the case-based system has robustness because they can always respond by modifying a similar case.

Several dialogue systems have been developed under the case-based approach. Murao et al. proposed a spoken dialogue system to provide shopping information to a person driving a car (Murao et al., 2003). To generate the response to an utterance, this system uses hand-annotated dialogue cases collected by the Wizard of OZ (WOZ) (Fraser and Gilbert, 1991) system in advance. Okamoto et al. proposed a dialogue agent for web guidance (Okamoto et al., 2001). This system is based on the WOZ method, but it is combined with case-based method for automatic response generation to reduce gradually the burden on the operator (wizard). The wizard checks each generated response and corrects it only when it is inappropriate. However, these systems cause the problem that manual operation is required. This means the advantage of the case-based approach (namely, low construction cost) is lost.

On the other hand, as a case-based system without hand control, a general-purpose chat system was proposed by Inui et al. (Inui et al., 2001; Inui et al., 2003). This system uses dialogue cases that are collected through all interactions with users and annotated automatically. The case is defined as a sequence of utterances and its response. However, their system involves the problems described below.

The first problem is that the similarity measure only depends on the information obtained from surface sentences of an utterance. As a result, the system can not distinguish two utterances that are the same sentence but have different intention. The meaning of an utterance changes according to how the word is expressed. For example, the response to the utterance "Pardon?" in a normal, puzzled, or angry manner should be just repeating the sentence, by saying it again with paraphrase, or by saying something different. In this way, natural human communication uses various modes of information. According to the published findings from psychological research (Mehrabian, 1972), only 7 percent of information is communicated verbally (through words), while the remaining 93 percent is communicated nonverbally (38 percent through the use of the voice, and 55 percent through facial expressions, body posture, gestures etc.). We believe nonverbal information is therefore necessary for dialogue systems to interpret the user's utterances more correctly.

The second problem is that the system's self-learning is only addition of the cases. For example, when the system tries to respond to "What's your name?", the following two past dialogues are put forward as similar cases:

**Case 1**:

A: "What's your name?"
B: "Today is my birthday."

**Case 2**:

A: "What's your name?"
B: "My name is Mary."

Although Case 1 is a system's inappropriate automatic response, Inui et al.'s system chooses it at a probability of 1/2. Moreover, if the inappropriate case is selected, another failed case (current generated dialogue) is added to the case-base, and increases the probability of a miss selection to 2/3 in the next selection. This is caused by the lack of learning mechanism of distinction between successful and failed cases.

The third problem is that case selection depends on only the similarity with current context, and the system does not care for the following turn. For example, there is a following two similar cases to the user's utterance "I lost my dear necklace":

**Case 3**:

A: "I lost my dear necklace."
B: "You're so careless."
A: "... Terrible!"

**Case 4**:

A: "I lost my dear necklace."
B: "That's too bad. Cheer up."
A: "Thank you."

Both Case 3 and Case 4 have the same utterance to user's input, and Inui et al.'s system chooses Case 3 at a probability of 1/2. However, the system's response in Case 3 angers the user, in comparison with comfort in Case 4. As shown in this example, it is important for case selection to consider the following user's reaction.

In light of the above-described problems, we propose a new method for a case-based natural language dialogue system. Although our system is based on the system proposed by Inui et al., it provides one solution to the problems described above by using a user's facial expressions that accompany their utterances. Our system uses the facial expressions for the following purposes:

- To improve the accuracy of similar case retrieval.
- To evaluate the appropriateness of similar cases for optimal case selection.
- To enhance the system's utterances to the user.

## 2 Case-based Dialogue System using Facial Expressions

The outline of our system is shown in Figure 1. In this system, a user and the system give utterances alternately, and one utterance consists of several sentences and one facial expression. When a user inputs one utterance, at first, the system extracts

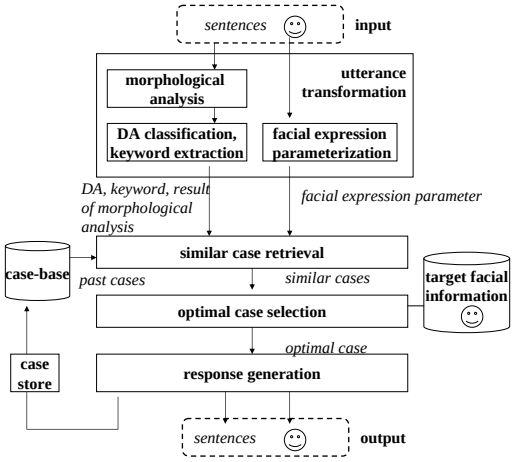



Figure 1: System overview

the linguistic and facial information. Secondly, the system considers the current context as a new problem, and selects similar cases by using linguistic and facial similarity measures. In the next step, the appropriateness of each similar case is evaluated by using facial information of user's reaction, and the optimal case is selected. Then, the selected case is adapted to the current context to generate the response to the user. Finally, the current user's input utterance and system's output utterance are added to the case-base.

### 2.1 Case Expressions

The case-base contains the time-series utterance history. The form of one utterance is as follows:

**ID Number**

- Utterance number

**Sentences**(For each sentence:)

- String
- Result of morphological analysis
- Dialogue act (DA)
- Keywords (a noun,a verb,an adjective)

**Facial expression**

- Parameters to represent a facial expression

Utterance number is a sequential serial number of the utterance, and a DA is a type of sentence indicating user's intention. Keywords are meaningful words indicating the topic of an utterance.

### 2.2 Utterance Transformation

When the user inputs one utterance, the utterance transformation module transforms it to the same form as with case expression.

First, the input sentences are divided into individual sentences. A morphological analyzer (Inui and Kotani, 1999) is used to analyze them into a series of words and parts of speech, and passes the results to the DA classifier (Inui et al., 2001) and keywords extractor (Inui et al., 2001). The DA classifier, trained from a DA-tagged corpus, determines a DA for each sentence. There are 17 types of DA, as listed in Table 1.

Table 1: Dialogue acts

| | | |
|---|---|---|
| *greet* | *request_comment* | *reject* |
| *bye* | *request (Y/N)* | *deliberate* |
| *opinion* | *confirm* | *apologize* |
| *will* | *request* | *surprise* |
| *explain_fact* | *suggest* | *thank* |
| *give_reason* | *accept* | |

Meanwhile, keyword extractor computes the weight of each word with heuristic rules which focus on "parts of speech", "kinds of characters" (*kanji*, *katakana*, *hiragana* in Japanese), "position in the sentence" (as a substitute for syntactic analysis), and so on. Then, a triplet of a noun, a verb, and an adjective is extracted as the keywords from each sentence.

The facial expression is represented by 18 parameters. There are 15 characteristic points on the eyebrows, eyes and mouth of the face, and the value of each parameter is given as the distance between two different characteristic points. The parameters and the characteristic points are shown in Figure 2. The nose has no characteristic points, since change of the facial expression hardly ever appears in the nose. An example result of an utterance transformation is shown in Figure 3.

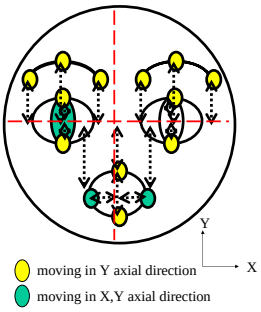
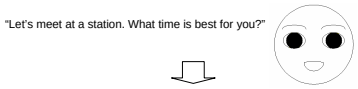



Figure 2: Facial characteristic points and parameters




| | |
|---|---|
| **Strings** | Let's meet at a station. What time is best for you?" |
| **Result of morphological analysis** | [Let_VM0 , 's_VM0, meet_VVI, at_PRP, a_AT0, station_NN.]<br>[What_DTQ, time_NN1, is_VBZ, best_AJS, for_PRP, you_PNP, ?] |
| **DA** | opinion, request_comment |
| **Keywords** | (station, meet, -),<br>(time , -, best) |
| **Parameters of facial expression** | P1 =1960,P2=2480, …, P18=4480 |

Figure 3: Example result of utterance transformation (original is in Japanese)

### 2.3 Similar Case Retrieval

The similar case retrieval module considers the sequence of the past $M$ utterances during the current dialogue as the current context, and selects similar cases in contrast to the sequence of the utterances in the case-base. In this paper, we set $M$ to two.

Throughout this paper, we represent the current context as $\mathbb{P} = \langle p_1, p_2 \rangle$, where $p_1$ is the last system's output and $p_2$ is the current user's input. On the other hand, a case is a sequence of time-series $J$ utterances in the case-base, and it is expressed as $\mathbb{C}_i = \langle c_i, c_{i+1}, c_{i+2}, ..., c_{i+J-1} \rangle$, where $c_i$ is an utterance with the utterance number $i$ in the case-base.

For calculating similarity between current context and each case in the case-base, we use the following three methods:

1. DA-based matching
2. Keyword matching
3. Calculation of facial expression similarity

The techniques that Inui et al. developed for DA-based matching and keyword matching are used in this module. DA-based matching is used for the case filtering based on a type of sentences. Keyword matching is the cost calculation based on the number of matched terms. Refer to (Inui et al., 2001) for further information. In this paper, the similarity between two sets of keywords, $s$ and $t$, is expressed as $Sim_{\text{key}}(s,t)$.

The similarity between two facial expressions, $x$ and $y$, is calculated from Formula (1) using 18 parameters of distance between the characteristic points.

$$Sim_{\text{face}}(x,y) = \sum_{i=1}^{18} W_i \frac{\{x[i]-y[i]\}^2}{\{\text{MAX}_i\}^2}, \quad (1)$$

where
$W_i$:weight of $i$-th parameter,
$\text{MAX}_i$:maximum value of $i$-th parameter,
$x[i]$ and $y[i]$:value of $i$-th parameters of facial expression $x$ and $y$ respectively

Then, the similarity between two utterances, $u$ and $v$, is calculated by using both similarity for keywords set and similarity for facial expression:

$$Sim(u,v) = \alpha Sim_{\text{key}}(key[u], key[v]) + \beta Sim_{\text{face}}(face[u], face[v]), \quad (2)$$

where
$key[u]$:a set of keywords of utterance $u$,
$face[u]$:facial expression of utterance $u$,
$\alpha, \beta$:constant values

The similar case retrieval module calculates the total similarity $Sim_{\text{ret}}(\mathbb{P}, \mathbb{C}_i)$ by using Formula (3), and retrieves $K$ most similar cases to the current context $\mathbb{P}$.

$$Sim_{\text{ret}}(\mathbb{P}, \mathbb{C}_i) = Sim(p_1, c_i) + Sim(p_2, c_{i+1}) \quad (3)$$

### 2.4 Optimal Case Selection

After similar cases have been retrieved from the case-base, the optimal case selection module selects an optimal case from them and uses it to generate the response to the user.

As mentioned in Section 1, it is important that a case-based dialogue system guesses the following dialogue and selects the case by measuring the appropriateness of the look-ahead section of each case in order to generate the appropriate response.

To measure the appropriateness, the user's feedback information about the quality of a system's response is useful. As feedback information, our system utilizes the facial expression of the $N$ utterances uttered right after the system's response. The number of $N$ utterances is fixed at one, that is, the system only uses the user's utterance uttered right after the system response. The case is, therefore, expressed as a quadruplet of the utterances.

The calculation of the appropriateness is explained informally as follows. As shown in Figure 4, after similar cases are obtained, utterance $c_3$ immediately following utterances $c_1$ and $c_2$, which are similar to the current context, is the candidate for the system's response. The facial expression of a utterance $c_4$ right after the candidate utterance $c_3$ is used as user feedback information, and the system compares it with the target facial expression. The appropriateness of case $\mathbb{C}_i = \langle c_i, c_{i+1}, c_{i+2}, c_{i+3} \rangle$ for the target facial expression $q$ is formally represented as the following Formula(4), by using the similarity for facial expression between $c_{i+3}$ and $q$.

$$App(\mathbb{C}_i, q) = \begin{cases} Sim_{\text{face}}(face[c_{i+3}], q) & \text{(if } q \text{ is desirable)} \\ 1 - Sim_{\text{face}}(face[c_{i+3}], q) & \text{(if } q \text{ is undesirable)} \end{cases} \quad (4)$$

As the target facial expression, either a desirable or an undesirable facial expression can be set. If we set a desirable target, the similar case has priority for selection; and if it is an undesirable target, the priority of similar case is low. We adopted a strategy of using a "smiling face" for a desirable target and an "angry face" for an undesirable target. However, desirable facial expressions will vary according to various factors, such as the domain of the system and the duration of a dialogue. We therefore presume that users can dynamically specify the target facial expression, and that more

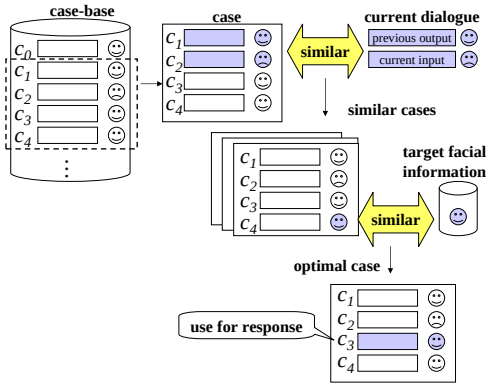



Figure 4: Optimal case selection

than one facial expression can be set and switched dynamically according to policy settings.

The optimal case selection module considers two factors comprehensively to choose one optimal case; one is the similarity to the current context, and another is the appropriateness of cases using the similarity to the target facial expression. The optimality between current context $\mathbb{P}$ and the case $\mathbb{C}_i$ is calculated as shown in Formula (5). Then, the case $\mathbb{C}_i$ which has a minimum score of $Opt(\mathbb{P}, \mathbb{C}_i)$ is chosen as the optimal case.

$$Opt(\mathbb{P}, \mathbb{C}_i) = \gamma Sim_{\text{ret}}(\mathbb{P}, \mathbb{C}_i) + \delta App(\mathbb{C}_i, q), \quad (5)$$

where

$\gamma$, $\delta$:constant values

### 2.5 Response Generation

After an optimal case is selected, the system uses third utterance in the quadruplet expression of the optimal case as a template for the response utterance. The adaptation of the template to the current context is done as follows. The output sentences are generated by replacing each keyword of the template with the corresponding keyword in the current context. To replace the keyword, we use the Inui et al.'s technique (Inui et al., 2001), which uses the keyword correspondence table made in keyword matching process, is applied. On the other hand, the facial expression of the optimal case can be directly used as the system's output.

### 2.6 Case Store

The case store module stores the pair of the user's input utterance and the system's output utterance in the case-base. As the dialogue is repeated, the input and output utterances are accumulated in the case-base in chronological order.

## 3 Empirical Evaluation

We made a prototype of the system for testing. Compared with Inui et al.'s system (Inui et al., 2001), the appropriateness of responses in our system was confirmed to be better. Some examples of actual dialogues that represent the advantage of our system over Inui et al.'s system are given in the following.

Dialogue example 1 (see Figure 5) shows the advantage of using facial expression information for similar case retrieval. Two input dialogues containing the same sentences but different facial information are considered. Although Inui et al.'s system generated the same response for these inputs, our system generates more appropriate responses according to the input facial information.

On the other hand, dialogue example 2 (see Figure 6) shows the advantage of optimal case selection. Case 1 and Case 2 are selected as similar cases to the current context, since utterances U1 and U2 are similar to those of the current context. Inui et al.'s system chooses Case1 as a similar case, although the user is angry in Case 1 because the system's response U3 is inappropriate. However, in our system, the facial expression shown in Figure 7 was set as an undesirable target in this experiment. The appropriateness of Case 1 is much lower than that of Case 2, because the similarity between the facial expression of U4 and the undesirable target facial expression is much higher. Therefore, overall, the system uses Case 2 to generate the responses shown in Figure 6.

## 4 Conclusion

A new method for case-based natural language dialogue system was developed. To generate an appropriate response, this system obtains the user's facial expressions and uses them to retrieve similar cases to the current context. Moreover, the system uses the user's facial information to evaluate the

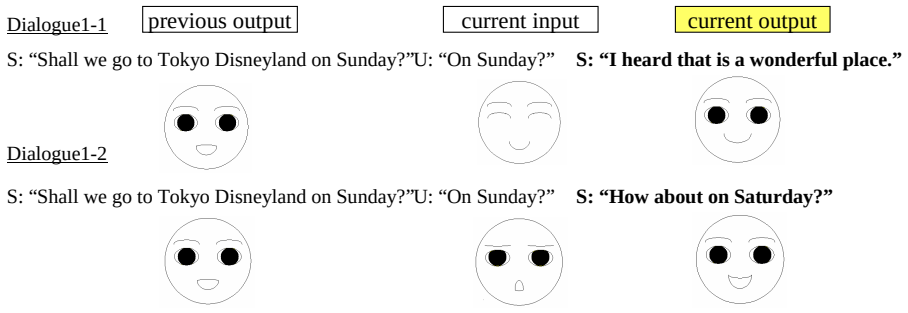



Figure 5: Dialogue example 1 (original dialogue is in Japanese)

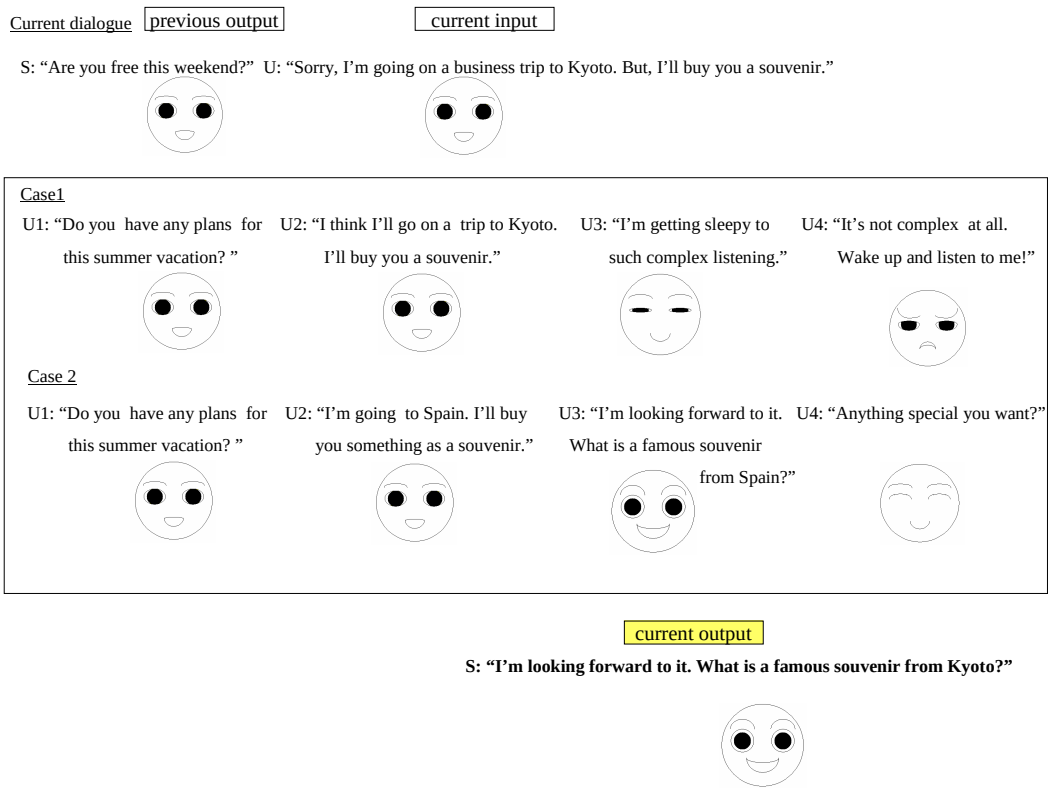



Figure 6: Dialogue example 2 (original dialogue is in Japanese)

Figure 7: Undesirable target facial expression

appropriateness of each case and to choose the optimal case. We plan to provide a more detail evaluation of our current system. After much experimentation, we would like to show the advantage of our system over other systems. We also plan to adopt an automated recognition technique of facial expression (Mase, 1991) to reduce the user's task because our current prototype system requires the user to input the facial expression manually.

## References


J. Allen, L. Schubert, and et al. 1994. The TRAINS project: A case study in building a conversational planning agent. TRAINS Technical Note 94-3, Univ. of Rochester.

S. Carberry. 1990. *Plan Recognition in Natural Language Dialogue*. The MIT Press, Cambridge MA.

N. Fraser and G. Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, 5(1):81–99.

N. Inui and Y. Kotani. 1999. Finding the best state for HMM morphological analyzer. *NLPRS'99*, pages 44–49.

N. Inui, T. Ebe, B. Indurkhya, and Y. Kotani. 2001. A case-based natural language dialogue system using dialogue act. *IEEE International Conference on Systems, Man, and Cybernetics*, pages 193–198.

N. Inui, T. Koiso, J. Nakamura, and Y. Kotani. 2003. Fully corpus-based natural language dialogue system. *2003 AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pages 58–64.

D. B. Leake. 1996. CBR in context:the present and future. In *Case-Based Reasoning - Experiences,Lessons, & Future Directions*, chapter 1, pages 3–30. AAAI Press & The MIT Press, Menlo Park California & Cambridge MA & London England.

K. Mase. 1991. Recognition of facial expressions for optical flow. *IEICE Trans. on Information Systems*, E-74(10):44–49.

A. Mehrabian. 1972. *Nonverbal Communiation*. Aldine-Atherton, Chicago.

H. Murao, N. Kawaguchi, and et al. 2003. Example-based spoken dialogue system using WOZ system log. *4th ACL SIGDIAL Workshop on Discourse and Dialogue(SIGDIAL-2003)*, pages 140–148.

M. Okamoto, Y. Yang, and T. Ishida. 2001. Wizard of Oz method for learning dialog agents. *International Workshop on Cooperative Information Agents (CIA-2001)*, pages 20–25.

J. Weizenbaum. 1966. ELIZA - a computer program for the study of natural language communications between men and machines. *CACM*, 9(1):3–45.