Understanding in Transformation of Written Directions into Maps

TEIJI FURUGORI*

A transformation of michi-annai-bun, written directions, into a michi-annai-zu, map, on a display device is described. It involves three distinct processes: sentence analysis, internal code generation, and map drawing. The sentence analysis establishes case frames for the incoming sentences; the internal code generation produces a sequence of matrices, each containing a motion or scene from the directions; and the map drawing computes a map by interpreting the contents of the matrices.

The maps drawn by the computer are as good as those drawn by humans. However, the model does not include a procedure that checks if michi-annai-bun's are coherent.

A complete understanding of a michi-annai-bun is achieved when the linguistic process is integrated with common sense knowledge, and spatial knowledge of the real world situation, i.e., when the map is followed.

1. Introduction

Spatial knowledge as a cognitive or mental map has been studied in various disciplines [1], [2], [3]. In the field of AI it relates to research in common-sense knowledge.

Kuipers [4] proposed representations of large-scale maps with a program that assimilated input sentences into a cognitive map representation. He then used it to solve route-finding and relative-position problems. Riesbeck [5] presented a program that judged the clarity and cruciality of the sentences in written directions. The research reported in this paper takes another approach to spatial understanding. We describe a program that transforms written directions into a universal second language, pictures or maps.

The transformation of written directions is interesting. Computationally its results show the extent of linguistic understanding in the pictorial forms. The lack of information in the transformation processes makes us consider the nature of understanding natural language and the requirements for a deeper, cognitive understanding.

2. Michi-Annai-Bun

In samples of michi-annai-bun's (literally "street-guide-sentence" in Japanese) written by forty students, we found three types of sentences: move-sentences, exist-sentences, and comment-sentences. A move-sentence describes the motion between two points. An exist-sentence shows the presence of landmarks on the route. A comment-sentence gives non-topographic information. Some examples of each of these sentences are:

Move-sentences

Go straight until you come to an intersection.
Turn right.

Exist-sentences

On your left you see a big department store.
There is a white house at the corner.

Comment-sentences

This place is dirty.
You can't miss it.

A complete michi-annai-bun is a sequence of these sentences. An example:

> Get off at Chofu station, Keio line, and take the north exit. Go for a short distance and you come to Kyukosyu-kaido. Turn right and you immediately see Tokai-bank on your left. Turn left on this side of the bank, go a while, and further, and then you come to the wide Kosyu-kaido. Turn left and then you will see the Dentsuu-dai on your right.
>
> [KEIO-SEN CHOFU-EKI-DE ORI, KITAGUCHI-NI DERU. SUKOSHI IKU-TO KYUKOSYU-KAIDO-NI TSUKIATARU-NODE MIGI-NI IKU-TO SUGU HIDARIGAWA-NI TOKAI-GINKOU-GA ARU. SONO TEMAE-NO MICHI-O HAITTE SHIBARAKU IKU-TO JUU-JIRO-GA ARU. SOKO-O SUGITE SARANI IKU-TO HIROI KOSYU-KAIDO-NI DERU. SOKO-O HIDARI-NI MAGATTE IKU-TO MIGIGAWA-NI DENTSUU-DAI-GA ARU.]

3. Transformation

We transform michi-annai-bun's into michi-annai-zu's (literally "street-guide-map") in three steps: sentence analysis, internal code generation, and map drawing [6].

*University of Electro-communications, Chofu, Tokyo, Japan.

Sentence Analysis

The michi-annai-bun is presented in romanized characters with spaces inserted between bunsetsu's (roughly, minimal phrases) to avoid the boundary problems in Japanese writing. The sentence "go further on the street and then . . ." is input as:

sono michi-o sarani massugu iki, sorekara . . .

Given such a michi-annai-bun, sentence analysis transforms it into a standard case-frame form in two steps:

(1) michi-annai-bun→simple sentences

(2) simple sentences→standard case-frames

The first decomposition identifies the verbs used in the michi-annai-bun.*[1] This uses a dictionary; for example, the word *iki* is identified in the dictionary as a move-verb whose root form is *iku* (to go).

The decomposition of a simple sentence into the standard form identifies the cases associated with the verb (V) of the sentence [7]. While more than thirty cases have been identified in an MT project [8], we find that five are adequate for move-sentences and two suffice for an exist-sentence. For the former the cases are origin (Or), goal (G), way (W), direction (Dr), and distance (Ds). For the latter, they are position (P) and object (Ob). Thus, the case frames for the michi-annai-bun take one of the two forms:

V (Or, G, W, Dr, Ds)

V (P, Ob)

Cases correspond to certain phrases in the sentence. The fillers for Ds, and in some instances for Dr, are derived from adverb phrases that contain words for distances or directions. For others, the case of a phrase is marked by the type of particle attached to the head noun of the phrase (See Table 1).

The particles are not always unambiguous indicators of case. The particle, *ni*, for example, can mark the three cases: G, Dr, and P. When this ambiguity exists, the identification of a case can sometimes be done by examining the type of the noun or verb used with the particle. Thus, the *ni*'s are marking the cases for P, Dr, and G, respectively, in *migi-ni aru* (to exist on the right), *migi-ni iku* (to turn to the right), and *michi-ni deru* (to come to a street).

Table 1 Typical verbs, cases, and case markers for michi-annai-bun's.

	Verbs	Cases	Markers
Move-sentences	*iku, susumu, mukau, naru, oreru, hairu, magaru, aruku,* etc.	Or	*kara, yori*
		G	*made, ni, e*
		W	*o, de*
		Dr	*ni, e,* (adverbs, e.g., *sarani*)
		Ds	(adverbs, e.g., *massugu*)
Exist-sentences	*aru, mieru, deru, desu, tsukiataru,* etc.	P	*ni, de*
		Ob	*ga, o*

*[1]A verb used immediately preceding an adjective or a noun is not considered as a verb. Thus, *aru* in *byoin-no aru kado-o magaru* (Turn the corner at a hospital.) is considered as part of a noun phrase. A similar example is seen in *migi-ni hairu michi* (the street going to the right). *Aru* and *hairu* are verbs in other instances.

In the portion of a michi-annai-bun shown earlier, the word *iki* (*iku* in its root form) is found as the verb for a move-sentence. The simple sentence thus is:

sono michi-o sarani massugu iku.

Using case markers and dictionary look-up, the case slots of the move-sentence are filled to produce:

iku (-, -, sono michi, sarani, massugu)

or

go (-, -, that street, further, straight)

Some sentences cannot be analysed by the procedure above. In these cases we check the sentence against an open list of sentential patterns and associated decompositions. This ELIZA-like provision [9] is a stop-gap measure to minimize failure due to incomplete linguistic coverage. If pattern matching fails, we treat the sentence as a comment-sentence and ignore it

The sentences

soko-ga boku-no ie desu. (That is my house.)

sugu wakarimasu. (You can find it easily.)

are unanalyzable. The first is considered as an exist-sentence, but the case P is not present. However, it matches the pattern:

X-ga Y-no Z-desu.⇨X-ni Y-no Z-ga aru.

where X, Y, and Z are arbitrary nouns. Therefore, it is decomposed according to the right hand side. The second sentence is ignored as a comment-sentence.

Embedded sentences pose another problem. The sentence

massugu migi-ni byoin-o mite susumu.

(Go straight while seeing a hospital on your right.)

consists of two simple sentences:

massugu susumu. (Go straight.)

migi-ni byoin-o miru. (You see a hospital on your right.)

This problem is solved, however, since most embedded sentences take a form in which an exist-sentence is embedded in a move-sentence; when we identify the cases for a move-sentence in an exist-sentence, we postulate that they are associated with the move-verb in the following (outer) sentence.

Fig. 1 shows the standard forms generated for the michi-annai-bun in the section MICHI-ANNAI-BUN. The verbs have been changed to their root forms, and case markers and any unnecessary conjunctions have been eliminated.

Internal Code Generation

This process analyzes each phrase assigned to the cases and transforms it into a matrix of one of three types: landmark, street, and motion.

We can get the properties of a street from a phrase. An adjective in a phrase may contain information (pro-

keio-sen chofu-eki-de oriru kitaguchi-ni deru.	→	eki(keio-sen,chofu,kitaguchi)*
sukoshi iku.	→	iku(-,-,-,-,sukoshi)
kyukoshu-kaido-ni tsukiataru.	→	tsukiataru(-,kyukoshu-kaido)
migi-ni iku.	→	iku(-,-,-,migi,-)
sugu hidarigawa-ni tokai-ginkou-ga aru.	→	aru(sugu hidarigawa,tokai)
sono temae-no michi-o hairu.	→	hairu(-,-,sono temaeno michi,-,-)
shibaraku iku.	→	iku(-,-,-,shibaraku,-)
juujiro-ga aru.	→	aru(-,juujiro)
soko-o sugiru.	→	sugiru(-,-,soko,-,-)
sara-ni iku.	→	iku(-,-,-,-,sarani)
hiroi kosyu-kaido-ni deru.	→	deru(-,hiroi kosyu-kaido)
soko-o hidari-ni magaru.	→	magaru(-,-,soko,hidari,-)
iku.	→	iku(-,-,-,-,-)
migigawa-ni dentsuudai-ga aru.	→	aru(migigawa,dentsuudai)

Fig. 1 Standard forms (for michi-annai-bun in the previous section).

*This is produced by a special routine that takes care of the situation at the train station.

perty) about a landmark. For example, from

shiroi ookina byoin-no aru kado-o magaru.

⇨magaru (-, -, shiroi ookina byoin-no aru kado, -, -)

we get the color (shiroi=white) and the size (ookina=big) of the type of the landmark (byoin=hospital).

A simple sentence produces one or more matrices, depending on the structures of the phrases in it. The matrices produced for the sentence above are shown in Fig. 2. In its full form each matrix includes the following properties:

STREET

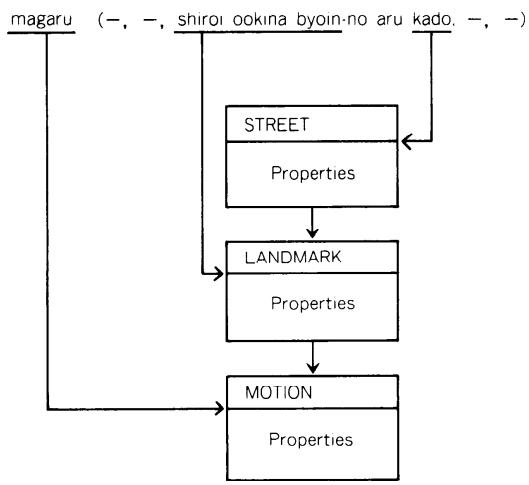


Fig. 2 Matrix generation and sequencing.

TYPE:	cross intersection, railway crossing, Y-intersection, etc.
WIDTH:	narrow, medium, or wide.
NAME:	(name of street)
LANDMARK	
TYPE:	school, hospital, tower, etc.
SIZE:	small, medium, or big.
COLOR:	white, red, etc.
POSITION:	right, left, front, etc.
NAME:	(name of landmark)
MOTION	
DIRECTION:	straight, right, left, etc.
DISTANCE:	ϕ, short, medium, long.

The phrases for the case slots are all considered to be noun phrases, except for direction and distance. Accordingly, the analysis of noun phrases as in Biermann and Ballard [10] is the key to the internal code generation.

Determining the semantic structure of a noun phrase begins with identifying the noun(s) for a landmark and/or a street. Then the properties for the noun are determined by constituents in the noun phrase.

When a noun phrase containing both landmark and street occurs, we make two assumptions to generate matrices: that the constituents preceding the landmark noun are associated with it and the constituents following it are associated with the street noun.*[2] We also

*[2]There are cases when this rule is violated. For instance, in the phrase *ginkou-no mae-no michi* (the street in front of the bank), *mae* (front) is associated syntactically with *michi* (street), but semantically with the *ginkou* (bank). In interpreting such a phrase, the position *mae* is associated with the landmark in the matrix.

assume that the noun for the landmark is semantically subordinate to the street noun. Thus, phrases such as

ginkou-no kado (the corner at the bank)

shiroi ookina byoin-no aru kado (the street corner at the big, white hospital)

Tatsumiya-to iu tatemono-no mieru tokoro (the place where the building named Tatsumiya is seen)

have essentially the same semantic construction, and each generates two matrices: one for the street and one for the landmark.

During the course of the code generation, it is important to note that more than the syntactic structure must be examined. This happens particularly when ordinals such as *tsugi* (next), *saisho* (first), and *n-banme* (*n*-th) are used in noun phrases. *Ookina michi* (a big street) and *san-banme-no michi* (the third street), for example, are syntactically similar, but the *san-banme* in the second phrase is an ordinal and cannot be taken as a simple property in the street (or landmark) matrix.

The use of an ordinal in a phrase is considered to be a short hand for avoiding repetition. The *san-banme-no michi* is equivalent to repeating the following sentences three times:

(*sukoshi*) *iku*. (Go (for a short distance).)

michi-ga-aru. ((You come to an intersection (street).)

Fig. 3 shows the sequence of the matrices produced for the michi-annai-bun in the section Michi-Annai-Bun. Here, the type of matrix is coded numerically: 1 for landmark, 2 for street, 3 for motion, and 4, a special case for the train station. The properties in each matrix are coded numerically or symbolically; a partial explanation is given in the figure.

Map Drawing

Each matrix should now have sufficient information about a motion or a scene along the way to the next (intermediate) destination. Maps are computed, from the matrices, by a Fortran program that uses GPSL, a set of subroutines for graphics.

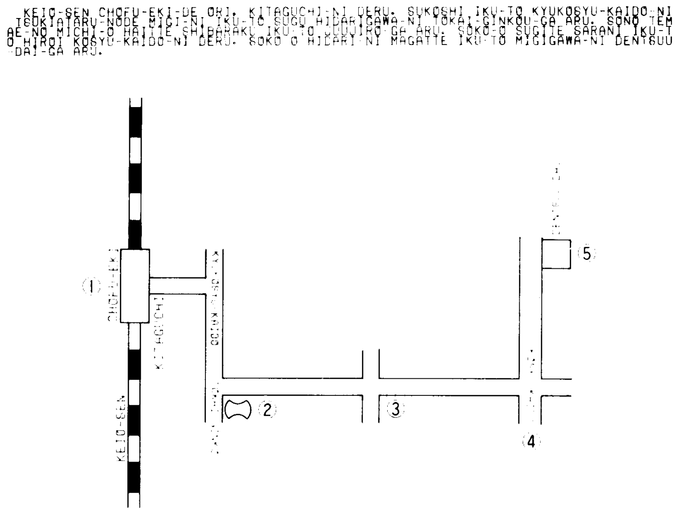
The Fortran program contains sets of subroutines for drawing intersections, landmarks, and motions. In addition, there is an important subroutine that controls the "you-are-here" pointer.

The "you-are-here" pointer is a structure that keeps track of the geographical (physical) space needed to draw a map. The variables used include:

PLACEX: current X coordinate on the plane
PLACEY: current Y coordinate on the plane
DIR: current direction
WIDTH: width of the street concerned

Once the sequence of matrices is produced, the map

Fig. 3 Sequence of matrices.



(1) CHOFU-EKI
(2) TOKAI-GINKOU
(3) SHIBARAKU IKU-TO JUUJIRO-GA ARU.
(Go for a while and you come to a crossroads.)
(4) HIROI KOSYU-KAIDO-NI DERU.
(You come to the wide Kosyu-kaido.)
(5) DENTSUU-DAI

Fig. 4

drawing is a simple, interpretive operation. Matrix 6 in Fig. 3, for instance, signifies a landmark with the special symbol on the map (BA for a bank), medium size, unidentified color, facing position left, and the name Tokai bank (TOKAI-GINKOU).

After the scene with the bank is drawn, the program interprets the subsequent matrices: a street with ⊣ shape, medium width, and unidentified name is drawn for matrix 7; matrix 8 produces a short, straight line on the same street; and so on. Finally, the interpretation of the matrices in Fig. 3 produces the map in Fig. 4 as the result of the transformation for the michi-annai-bun (in the section Michi-Annai-Bun).

4. Results and Limitations

Figs. 5 and 6 show typical michi-annai-bun's with the results of the transformations, michi-annai-zu's, shown as type (c). We add two sets of maps to each figure for further thoughts on understanding michi-annai-bun's: the (a) maps are drawn by the students who wrote the michi-annai-bun's; (b)'s are drawn from the michi-annai-bun's by students who had no knowledge of the places.

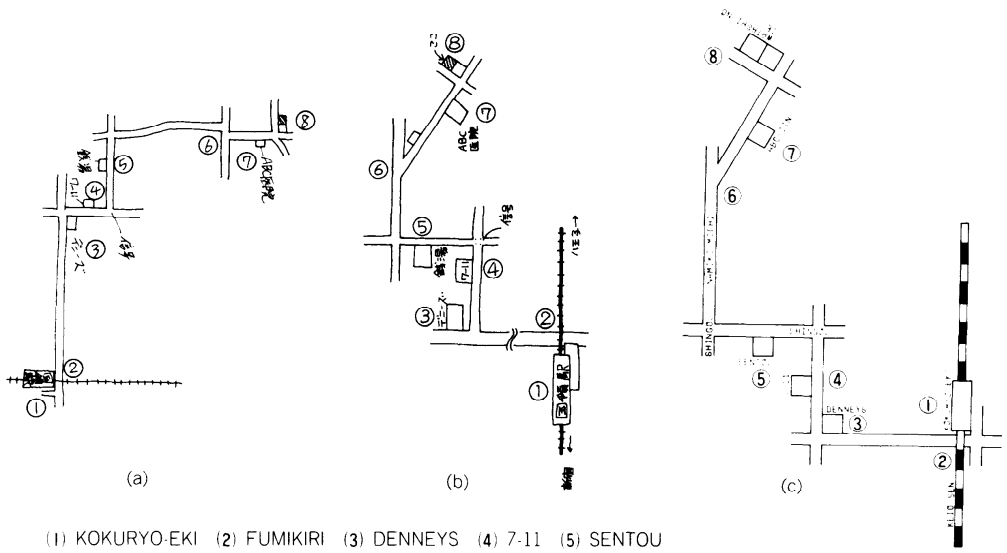
Sentences in michi-annai-bun's are short in length and both syntactically and semantically limited in scope. Technically speaking, however, there are sentences and phrases that are unpicturable and beyond the grasp of the present analytical procedures.

ginkou-no migi-to iu-yori naname-no michi-o iku.
(Take the street going by the bank diagonally than to the right.)
senro-to heikou-ni Kawagoe-homen-ni aruku.
(Walk toward Kawagoe along the railroad.)
kono hen-de kodomo-ga ippai ason-de imasu.
(There are a lot of kids playing aroung here.)
kono mise-de Yamada-no ie-o kiite kudasai.
(Ask Yamada's at this store.)
hiroi michi-ni deru-ga soko-made iku-to dame-desu.
((You) come to a wide street, but when you come to this point, it's no good (you have come too far).)
ookina michi-ni demasu-ga soremade-ni kousaten-ga muttsu aru.
(You will come to a big street, but before then you pass six intersections.)

The phrasal structures in the first pair and the sentential patterns in the second are not identifiable in the present transformation processes. The unpicturability of the third pair needs some explanation.

The first sentence in the third pair contains three simple sentences: *hiroi michi-ni deru* (come to a wide street), *soko-made iku* (come to that point), and *dame-desu* (it's no good). *Dame-desu* is then ignored as a comment-sentence and a partial map is drawn for the first two simple sentences. This process is wrong, however, because *dame-desu* negates the previous sentences. Similarly in the second sentence, the intersections are drawn incorrectly since the simple sentences are inter-

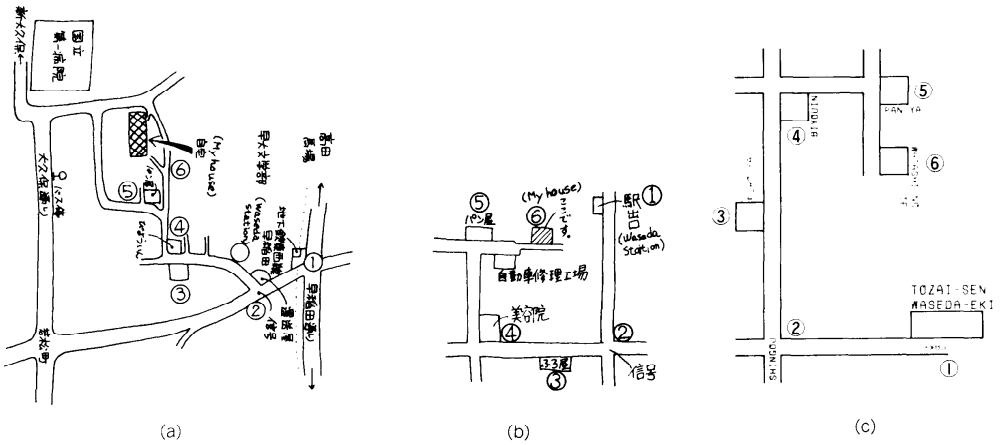
KEIO-SEN KOKURYO-EKI-NI GESHA. SUGU FUMIKIRI-O WATARI, HITASURA ARUKU. MIGITE-NI DENNEYS-GA MIETARA MIGI-NI IKU. SUKOSHI IKU-TO HIDARI-NI 7-11-GA ARI. SAISHO-NO SHINGOU-O HIDARI-NI IKU. HIDARI-NI SENTOU-GA MIETE, SUKOSHI IKI, SAISHO-NO SHINGOU-O MIGI-NI IKU. NAMIKI-MICHI-O HITASURA SUSUMI, SHINGOU-NO ARU KADO-O MIGI NANAME-NI IKU. SUKOSHI IKU-TO MIGI-NI ABC-IIN-GA ATTARA, SUGU HIDARI-NI IKI, MIGIGAWA-NO 2-KENME-NI WATASHI-NO IE-GA ARIMASU.



(1) KOKURYO-EKI (2) FUMIKIRI (3) DENNEYS (4) 7-11 (5) SENTOU
(6) KADO-O MIGI NANAME-NI IKU.
(Pass through the corner diagonally to the right.)
(7) ABC-IIN (8) WATASHI-NO IE (My house)

Fig. 5

CHIKATETSU TOZAI-SEN WASEDA-EKI-DE ORI, MIGIIE-NO SAKA-O NOBORU. 1-TSUME-NO SHINGOU-O MIGI-NI MAGARI, MASSUGU IKU-TO HIDARITE-NI FURO-YA-GA MIE, SARANI IKI, BIYOUIN-NO ARU KADO-O MIGI-NI MAGARU. SUKOSHI IKU-TO SHOUMEN-NI PAN-YA-GA ARI, SOKO-O MIGI-NI IKU. SUKOSHI IKU-TO HIDARIGAWA-NI WATASHI-NO IE-GA ARU.



(1) WASEDA-EKI (2) 1-TSUME-NO SHINGOU-O MIGI-NI MAGARU.
(Turn right at the first traffic lights.)
(3) FURO-YA (4) BIYOUIN (5) SHOUMEN-NI PAN-YA-GA ARU.
(There is a bakery in front of you.)
(6) WATASHI NO IE

Fig. 6

preted sequentially in the left-to-right fashion. We need some intersentential analysis to take care of these problems.

5. Implications

Understanding michi-annai-bun's appears easier than many other kinds of discourse. Their sentences are in most cases independently interpretable without paying attention to intersentential relations, e.g., casual links. The hierarchical knowledge structure, that is crucial for understanding a story [11], is not present.

Conjunctions and pronouns are important in text understanding [12], but they are uninteresting in michi-annai-bun's and considered to be "noise" or skippable elements [13], [14]. In the following examples, the conjunctions are skipped. The pronouns used refer to the place in the "you-are-here" pointer; the place in it works, most of the time, as the default value for pronoun references.

sukoshi itte *sorekara* migi-ni magaru.
(Go for a short distance, *and then* turn right.)
byoin-ga aru. *soko-o* migi-ni magaru.
(There is a hospital. Turn right *at the place.*)

Compared with the (a) maps in Figs. 5 and 6, the (b) and (c) maps are distorted in a number of places. The (b) and (c) maps, on the other hand, are strikingly similar.

The michi-annai-bun's could be incomplete and this may have caused the distortions. However, the students who drew the (b) maps from them indicated that they were perfectly clear and understandable. We hypothesize that michi-annai-bun's are understood initially without much spatial reasoning.

Consider the following sentences in Fig. 5.
sugu fumikiri-o wataru.
(Pass the railroad crossing immediately.)
Denneys-ga mietara migi-ni iku.
(When you see Denney's turn right.)

At first we are interested in recognizing actions or states with their proper participants, and do not care about the exact location of the railroad crossing or Denney's. The information is expected and supplied when we actually see such landmarks.

We must initially decide if the sentences satisfy conditions necessary for describing directions. At this time, we understand generally how the individual concepts such as *fumikiri* (railroad crossing) or *magaru* (to turn) are pictured, but we are not necessarily sure of their position, distance, angles, or other physical attributes.

Expressions such as *sukoshi* (a little), *magaru* (to turn), *kado* (corner), etc. are not considered conceptually ambiguous, but they are when drawing maps. Consider the discrepancies for *fumikiri* and *Denneys* in Fig. 5(b), and the angles of the intersections at *hitotsu-meno shingou* (the first traffic lights) and *kado-o migi-ni magaru* (turn right at the corner) in Fig. 6.

The similarity between (b)'s and (c)'s does not prove that the computer can understand michi-annai-bun's as well as we human beings do as the model does not include checks for their clarity and understandability. It shows however that, once this test is done, the computer is able to draw maps comparable to those produced by human beings.

Understanding michi-annai-bun's uses linguistic analysis, spatial knowledge, and perceptual information. It starts with the analysis of the sentences using knowledge structures (frames or scripts) to find the elements of directions. This process omits a lot of spatial reasoning. We human beings only do partial understanding initially [15], and perform the spatial reasoning later at "execution time", i.e., when following the directions. The complete understanding of michi-annai-bun's is thus completed when the knowledge structure becomes "filled" with visual information.

6. Summary

The experiment described here shows some interesting points in the relationship between the nature of natural language and its understanding.

We see the difference of the expressive power in natural language and pictures. Some concepts are expressed better in natural language and others in pictures. Local environments, relative positions and distances among places are usually better visualized than verbalized. The reverse is true of concepts such as "second or third street", "noisy place", and "the place with a good smell of coffee".

Linguistic understanding is a partial process of understanding something in the real world. The experiment shows that little or no spatial reasoning is involved initially in understanding michi-annai-bun's.

A complete understanding of a michi-annai-bun becomes possible when linguistic analysis is associated with common-sense knowledge and perception. However, the structure of the common-sense knowledge here could be quite dependent on the environment in which one lives [16].

Acknowledgements

The author is grateful to the anonymous referee(s) for insightful comments and suggestions. Thanks are also extended to Professors B. Phillips and W. R. McKinzie. H. Niitani and N. Hoshimoto gave assistance in programming.

References

1. DOWNS, R. M. and STEA, D. (eds.) Image and Environment, Aldine Pub. Co., Chicago, 1973.
2. PIAGET, J. and INHELDER, B. The Child's Conception of Space, Norton, New York, 1967.
3. LYNCH, K. The Image of the City, MIT Press, Cambridge, Mass., 1960.
4. KUIPERS, B. Modelling Spatial Knowledge, *Cognitive Science* 2, 2 (1978).

5. RIESBECK, C. K. You can't miss it: Judging the Clarity of Directions, *Cognitive Science* **4**, 3 (1980).
6. FURUGORI, T. Computing a Map from Michi-annai-bun or Written Directions *IJCAI-7*, Vancouver (1981).
7. FILLMORE, C. The case for case, In BACH, E. and HARMS, R. T. (eds.) *Universals in Linguistic Theory*, Holt, New York, 1968.
8. NAGAO, M. Outline of the Machine Translation Project at Kagaku Gijutsu-cho, 38th Workshop on Natural Language Processing, IPSJ, July 15, 1983.
9. WEIZENBAUM, J. ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine, *Comm. ACM* **9**, 1 (1966).
10. BIERMANN, A. W. and BALLARD, B. W. Toward Natural Language Computation, *American J. of Comp. Linguistics* **6**, 2 (1980).
11. RUMELHART, D. E. Notes on a Schema for Stories, In BOBROW, D. G. and COLLINS, A. (eds.), Representation and Understanding, Academic Press, 1975.
12. VAN DIJK, T. A. Text and Context, Longman, London, 1977.
13. SCHANK, R. C. et al. An Integrated Understander, *American J. of Comp. Linguistics* **6**, 1 (1980).
14. LEBOWITZ, M. Cancelled due to Lack of Interest, *IJCAI-7*, Vancouver (1981).
15. DEJONG, G. Skimming Stories in Real Time, Research Rep. 158, Dept. of Computer Science, Yale Univ. (1979).
16. SIMON, H. A. Cognitive Science: The Newest Science of the Artificial, In NORMAN, D. A. (ed.), Perspectives on Cognitive Science, Ablex, New Jersey, 1981.

(Received September 13, 1983; revised September 20, 1984)