A Connectionist Simulation of the Form-Meaning Relationship in Japanese Business Manuals

Yoichi Masuzawa

Abstract

For improving natural/computer language conversion algorisms, this paper proposes an account of the acquisition of semantic relations (form-meaning relationship) of Japanese complex sentences using the basic concepts of connectionism and a construction-based theory of grammar (Fillmore 1988), especially those parameters extensively asserted in Ohori(2001).

Firstly, this paper proposes that *semantic* relation (form-meaning relationship) emerges in *causal chain* (Croft 1991) and tries to map each complex sentence to semantic relations (i.e., *cause* and *reason*). The motivation behind the network is the notion that merely the drive to map an input sequence of sentences to a semantic relations-output is sufficient to induce the necessary internal abstractions to facilitate the mapping.

To test this hypothesis, a Simple Recurrent Network (Elman 1990) has been created and tested. This paper demonstrates the emergence of the notion of "meaning" in the neural network that successfully learns to map from "sequence of sentences (form)" to "semantic relations (meaning)".

KEYWORDS: NEURAL NETWORK, CONNECTIONIST SIMULATION, INFORMATION SYSTEMS, MACHINE TRANSLATION

1. INTRODUCTION

A natural/computer language conversion software, *Structured Business Manuals Analysis and Design Method for Management Information Systems* (SMAD: Masuzawa 2000), has been applied to business manuals written in natural languages (Japanese and English) of both Japanese and American companies and proven to be valid in terms of *simple* sentence translation to each equivalent systems diagram for designing information systems. Simple sentences, however, occupy less than 30% of sentences in the business manuals. As SMAD has been developed as a natural language (business manuals)/ computer language (systems analysis diagrams and computer programming language) conversion tool, not only *simple* sentences (one-verb sentences) but also *complex* sentences (two-verb sentences) should effectively be processed and converted using the basic concepts of connectionism and cognitive linguistics theories.

As Diessel(2002) mentions, many previous accounts of complex sentence processing and its acquisition using traditional connectionist models focus their concentration on *grammatical relations* (subject, object etc.: Morris et al. 2000), syntax (Elman 1990), and the importance of starting *small* (i.e. simple sentences) (Elman 1990). This focus is necessary to provide an assumption that the preceding connectionist simulator can only process *randomly* generated sequential data that is trying to map a sequence of "words" to *grammatical relations*, not to *semantic relations* (*cause, reason* etc.: Ohori 2000) which are crucial in terms of effectively and precisely analyzing and designing business manuals and/or management information systems.

In contrast to this approach, this paper proposes that *semantic* relation (form-meaning relationship) emerges in *causal chain* (sequence of sentences which are semantically related: Croft 1991) rather than in randomly generated sentences and tries to map each sentence to semantic relation (i.e. *cause, reason, temporal sequence* and others). The motivation behind the network is the hypothesis that merely the drive to map an input sequence of sentences to a semantic relations output is sufficient to induce the necessary internal abstractions to facilitate the mapping.

To test this hypothesis, a modified Simple Recurrent Network (Elman 1990) has been created and tested. This paper demonstrates the emergence of the notion of "meaning" in the neural network that successfully learns to map from "sequence of sentences (form)" to "semantic relations (meaning)". This paper also analyzes the hidden layer representations of the emergent semantic relations, and demonstrates that these representations correspond to the Lakoff's radial category.

2. SEMANTIC RELATIONS

Comparing the two sentences "I changed jobs and the salary was low" and "The salary was low and I changed jobs," the semantic relations represented by "and" are different from each other, as CAUSE relation in the former sentence and REASON relation in the latter. Thus the clause (verb) chain linked by "and" connector implies a semantic relation in the chain (ie. the meaning of the complex sentence). Roughly speaking, when the first verb denotes action and the second non-action, the action is considered to be "the cause of the second non-action," and the first non-action/ the second action combination makes the first non-action "the reason of the second action." The sentences, therefore, could be reworded as "My changing jobs made my salary low" and "Because my salary was low, I changed jobs" respectively.

The two sentences above can directly be translated into Japanese using "TE(=and)" connector with exactly the same mechanism of forming semantic relations (see Table 1). This means there are two types of "TE," : TE-Cause and TE-Reason. Those clause-linking structures linked by the TE connector are called "TE-form" which are the most frequently used constructions in business manuals. Roughly 40%-50% of Japanese sentences in business manuals include this "TE-form".

Semantic Relation	CAUSE	REASON
Causal Chain	Cause[ACTION]->Result[non- ACTION(state)]	Reason[non-ACTION(state)]->Result[ACTION]
Sample Sentence	Tensyoku si-**te** kyuuryoo ga yasui. Changing.a.job did-**TE** salary NOM be.cheap ACTION STATE	Kyuuryoo ga yasuku -**te** tensyoku sita. Salary NOM be.cheap-**TE** changing.a.job did STATE ACTION
Machine Translation	I changed jobs **and** the salary was low.	The salary was low **and** I changed jobs.

Table 1 CAUSE and REASON relationship and syntax

Note: NOM/Nominative case marker, yasuku- = yasui, si- = si-ta(past tense suffix).

It should be noted here that the semantic relations of two clauses linked by TE cannot be inferable from the meaning of individual words or constituents alone but the TE-linkage's (TE-form's) property. In other words, the construction (ie. form and meaning pair) defines the semantic relations. Normally, a construction-based theory of grammar (Fillmore 1988) with those parameters (*control, temporal sequence,* *action coherence* etc.) asserted in Ohori(2001) should be utilized extensively for analyzing semantic relations. In this sense, the stated cases above are straightforward and rather artificial.

For years, the machine translation theories avoided this issue in a clever way utilizing two ambiguous connectors such as "TE" and "and" as shown in Table 1. On the contrary, natural language/computer language conversion theory definitely requires strict algorism which clarifies the semantic relations between two clauses. Otherwise one cannot define correct data flows for management information systems.

So, this paper tried to utilize the neural network which simulates the human cognitive process for language acquisition in order to find out whether computers can automatically and efficiently understand the semantic relations of the "TE-form".

3. THE NETWORK ARCHITECTURE

Elman(1990) described a simple recurrent network that succeeded in assigning words to grammatical categories (such as noun and verb) on the basis of distributed evidence extracted from strings of words which followed a set of grammatical rules. The task involved presenting the network with a sequence of words (sentence), one at a time, and training it to output a prediction of the next word. There is no semantic information indicating meaningful relationships between words. The order of a sentence is random.

Our network architecture is also a simple recurrent network based on Elman's works. Unlike the original network, however, this paper tried to find out whether the network could successfully assign sentences to semantic relation categories (such as *cause* and *reason*). Therefore, the data (sentence) sequence was not randomly created but in the logical order (such as "TE-CAUSE, TE-REASON, …,"). For example, "Tensyoku si-te kyuuryoo ga yasui. Kyuuryoo ga yasuku-te tensyoku sita. (I changed jobs and the salary was low. The salary was low and I changed jobs (again).)."

Figure 1 shows the recurrent network used in this paper in which context units receive input from a layer of hidden units. The activities of the context units are direct copies of the hidden unit activities. Context units provide the network with a dynamic memory for learning clausal chains.

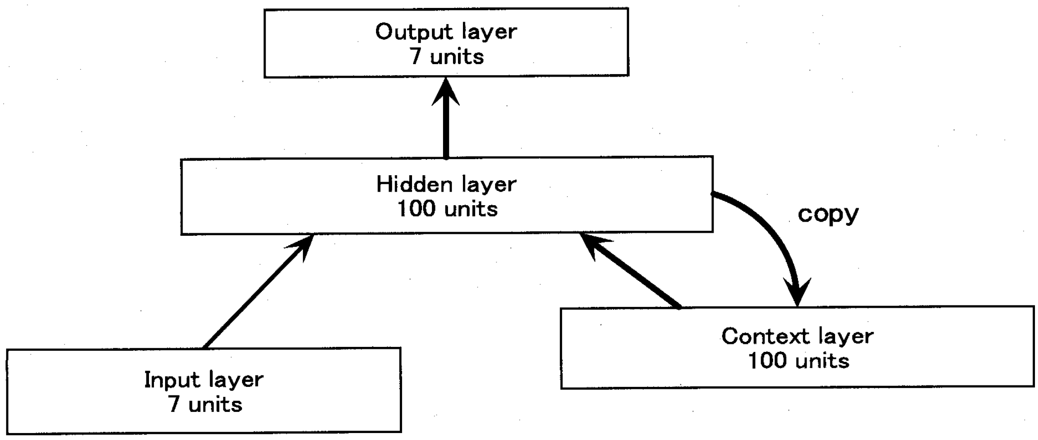

Fig. 1 Simple Recurrent Network used in this paper

Using the back propagation learning procedure, the network is taught to output a prediction of the next word. For example, if a word "Tensyoku " in the sentence "Tensyoku o si-te kyuuryoo ga yasui." is fed to the Input layer, the network is taught to output "o" to the Output layer.

The input vocabulary consists of 96 words (plus end of sentence marker), represented as 7-bit patterns, for example, "Tensyoku (changing a job)" is encoded as "1 0 0 1 1 0 1." Of these 96 words, 49 are predicates (41 verbs and 8 adjectives), 39 are nouns and remaining 8 words are a variety of function words. Nouns are divided into categories of persons (I, you, a friend, etc.), places (school, department store, etc.) or things (car, shoes, money, etc.). Of the verbs, 20 are action transitive/intransitive (buy, build, go, enter, etc.) and 21 are non-action intransitive and passive or middle voice(be able to meet, be pleased, sell in "This book sells well"). The 8 remaining words are 4 case markers (wa: topical, ga: nominatives, o: accusative, ni: accusative/locative), 1 demonstrative (this), 1 adverb (well), 1 past tense marker" ta," and the connecter "te."

The data corpus (583 sentences/4664 words) was composed of 312 *cause* relation sentences (55.4%) and 271 *reason* relation sentences (44.6%). Each sentence contained two verbs and one connector "te."

4. RESULTS

4-1 CAUSE and REASON RELATION

The network was trained and tested by *Tlearn* simulator (Plunkett and Elman 1997). Training (to output a prediction of the next word of the input word) was conducted for 107 epochs (500,000 sweeps for 4664 data), with 583 sentences in each epoch. The learning rate was 0.1, initial weight set within a range of 0.1. There was no momentum.

The type of predictions the network made is illuminating. Rather than the precise word, the network predicts the right grammatical and semantic relation category of the word. In other words, the network is capable of inducing the grammatical structure of the input sentence from the input strings themselves without any prior knowledge of the type of grammar that was used to set the training set.

For example, instead of producing "Tensyoku o site[ACTION] kyuuryoo ga yasui[non-ACTION]," the network creates "Benkyo o site[ACTION] minna ga yorokobu[non-ACTION]" which means "Everybody learned (it) and was pleased," both sentences are equal in terms of semantic relation type (i.e. CAUSE relation).

How a network represents individual words can be investigated using a cluster analysis and a principal component analysis (PCA). By averaging the hidden unit activities produced by individual words in many different sentence contexts, it is possible to investigate how different words cluster together (cluster analysis) and group together (PCA) forming a radial category of Lakoff.

Figure 2 shows the results of a cluster analysis (Figure 3 and 4 for PCA). The fact that the TE-Reason cluster (group) and TE-Cause cluster (group) are not on the same branch in the cluster analysis (not in the same group in PCA) means the network successfully learned semantic relations.

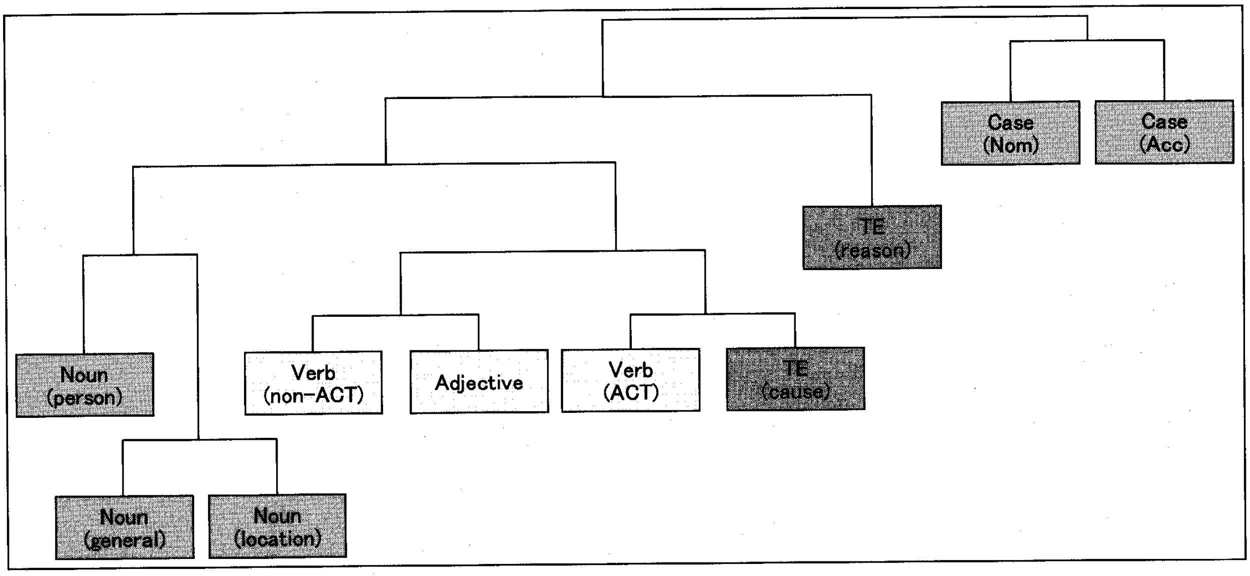

Fig. 2 A Cluster Analysis of Hidden Unit Activities (Simplified)

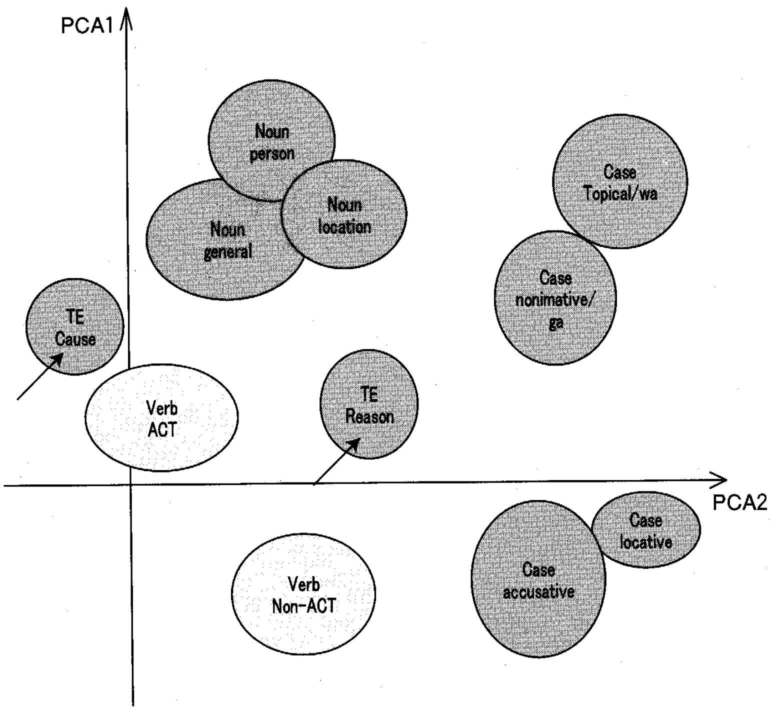

Fig.3 Principle Component Analysis (PCA:2Dimension:Simplified)

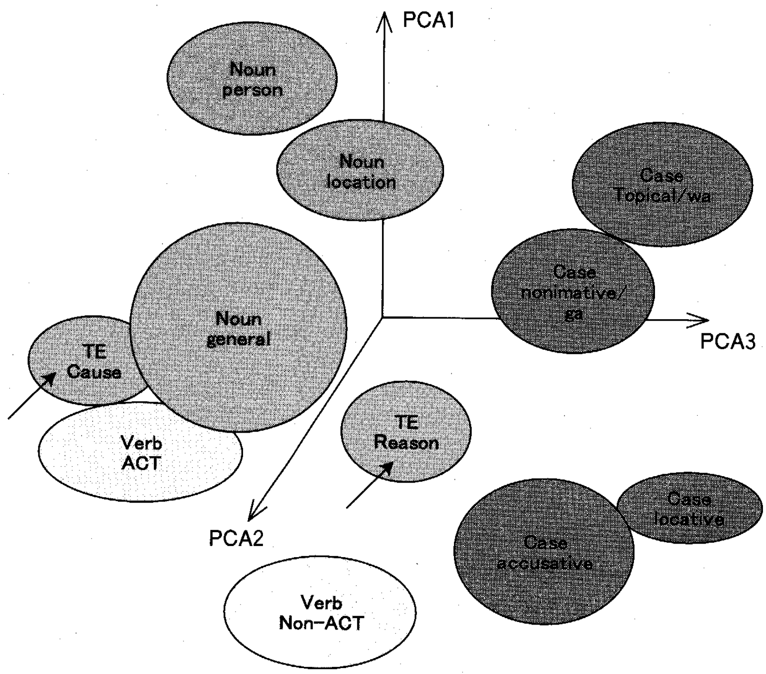

Fig.4 PCA (3Dimension:Simplified)

4-2 Non-Typical Relation (INTENTION)

If you look closely at Figures 3 and 4, you will notice that the grouping circles in terms of TE-CAUSE and TE REASON cover a rather wide range. This means there might be typical (ideal) CAUSE/REASON linkage and non-typical linkage. For example the network automatically concludes that following (1), (2) is much more typical than (3):

(1) Kyuuryoo ga yasuku-te tenshyoku sita. (EVENT(uncontrollable)+ACTION)=REASON
Salary NOM be.cheap-TE changing.a.job did
STATE ACTION
I changed jobs, because my salary was low.

(2) Tomodati ga waruku-te tensyoku sita. (STATE(uncontrollable) + ACTION) =REASON
a.friend NOM be.bad-TE (I)changing.a.job did
STATE ACTION
I changed jobs, because my friend was bad (guy).

(3) Tomodati wa waruku-te tensyoku sita. (STATE(controllable) + ACTION) = INTENTION
a.friend TOP be.bad-TE (he/she) changing.a.job did
STATE ACTION
My friend was bad (in attitude) and he changed jobs.

Ohori (2001) parameter explains these non-typical cases well using the [+control] feature. Ohori 2001 asserts that one way to systematize an unstructured set of semantic categories is to introduce decomposition. For example, we can introduce a feature like [+/- control], which may draw a line between INTENTION and REASON. Also, a feature like [+/- temporal overlap (proximity)] would be useful to systematize the semantics hierarchy (INTENTION-REASON-CAUSE…). In this sense, above (1) and (2) have uncontrollable V1 while V2 is controllable, i.e. [-control] for the former and [+control] for the latter. The network obviously could understand the discrepancy.

The network also learned from the input corpus that the subjects of first verb (V1) and the second (V2) in (3) are equal so that it could identify the difference between (1), (2) and (3) in terms of clause connection types (mixture of nexus and juncture). It should also be noted that the morphological difference between (2) and (3) is small (just a subtle case maker *ga* and *wa*, which are predicate (narrow) focus and sentence (wide) focus respectively).

In theory, the nexus can be divided into three types in relation to two parameters which are dependency and embedding (i.e. coordination:[+dependent, -embedded], cosub- (trans) ordination:[+dependent, -embedded] and subordination:[+dependent, +embedded]). In Ohori (2001), the juncture types are clearly defined as schema based decomposition as follows:

**Nuclear juncture*

The first category of inter-clausal semantic relations: atomic predicates (states or activities) plus their elaboration by addition of aspectual, causative, or directional operators.

**Core juncture*

The second category of inter-clausal semantic relations: units of first stage elaboration linked together beyond logical structure representation (e.g. CAUSE operator), but with obligatory participant sharing.

**Clausal juncture*

The third category of inter-clausal semantic relations: units of second stage elaboration linked together each having autonomy, characterized by non-obligatoriness of participant sharing and availability of separate perspectives (i.e. Anchoring points).

In pursuant to the above theory, the former (1),(2) are supposed to be clause-cosub (trans) ordination and the latter (3) core-cosubordination (transordination). (Ohori 2001).

The network judges the latter has stronger syntactic and semantic connection (CORE-COSUB(TRANS)ORDINATION:INTENTION) than the former (CLAUSE-COSUB (TRANS) ORDINATION:REASON) has, and this is supported by the theory as well.

4-3 Non-Typical Relations (INDEPENDENT and SETTING)

Other cases, namely INDEPENDENT EVENTS and SETTING, which are also successfully distinguished by the network as "non-typical" samples, are also explicable using Ohori parameters.

(4) Yo ga huke-te tori ga naku. (INDEPENDENT)
night NOM become.dark. birds NOM be.singing
non-ACTION **ACTION**
It gets dark and birds are singing.

(5) Sonoba ni tui-te huukei ga hirogatta. (SETTING)
there LOCATIVE arrive-TE scenery NOM wide-open.PST
ACTION **non-ACTION**
Arriving there, looked around and found unbounded vastness.

Normally, the network identifies (4) as REASON (non-ACTION+ACTION) and (5) as CAUSE (ACTION + non-ACTION) but the answer was no. The Ohori parameter set judges both sentences as INDEPENDENT and SETTING respectively ([-action coherence] [-temporal proximity]). This is strikingly matched with the output of the neural network.

The summary of the Ohori parameter combination with example sentences is as follows (excerpt from Ohori 2001):

(6) [+/- control] to distinguish between Purpose/Jussive and Direct perception.
(7) [+ action coherence] [+ temporal proximity]
I went into the room and searched for her trace.
(8) [- action coherence] [+ temporal proximity]
I went into the room and there was a time-bomb ticking.
(9) [- action coherence] [- temporal proximity]
I went into the room though there was no reason to do so.

5. DISCUSSION and CONCLUSION

The simulation was intended to demonstrate that the semantic relations of complex sentences are learnable. As is clear from the examination of the hidden layer, we can see how the network successfully stores the abstraction of the semantic relation.

We have been preceded by many others in the use of Simple Recurrent Network. Most of these previous works focused on the simple sentence and its grammatical relation. This paper proposes that *semantic* relation (form-meaning relationship) emerge rather in *causal chain* (sequence of sentences which are semantically related: Croft 1991) and tries to map a sequence of "sentences" to semantic relations. We succeed in two semantic relations, *cause* and *reason*. We also analyze the hidden layer representations of the emergent semantic relations and demonstrate that these representations correspond to a radial category of Lakoff.

There are, however, other semantic relations to be investigated in terms of TE form, such as *manner, means, material, measure, intend, contrast, setting* and so forth. Further research which utilizes those parameters (*control, temporal sequence, action coherence* etc.) extensively asserted in Ohori(2001) is required for accomplishing the analysis.

6. REFERENCES

1. Ohori, T. (2001). Some Thoughts on a New Systematization of Interclausal Semantic Relations, a paper read at Role and Reference Grammar Workshop, University of California, Santa Barbara: 1-6.

2. Masuzawa, Y. (2000). On Structured Manuals Analysis and Design for Japanese

Enterprises, Proceedings of AMCIS, Americas Conference on Information Systems 2000. pp.1421-1427.

3. Plunkett, K. and Elman, J. (1997). Exercises in Rethinking Innateness. The MIT Press, Boston.

4. Elman, J. (1990). Finding structure in time. Cognitive Science 14, pp. 179-211.

5. Diessel, H. (2002).The Acquisition of Complex Sentences. Cambridge University Press, Cambridge, United Kingdom.

6. Croft, W. (1991). Syntactic Categories and Grammatical Relations. The University of Chicago Press, Chicago.