Translation by Abduction

Jerry R. Hobbs and Megumi Kameyama
Artificial Intelligence Center
SRI International
Menlo Park, California

Machine Translation and World Knowledge. Many existing approaches to machine translation take for granted that the information presented in the output is found somewhere in the input, and, moreover, that such information should be expressed at a single representational level, say, in terms of the parse trees or of "semantic" assertions. Languages, however, not only express the equivalent information by drastically different linguistic means, but also often disagree in what distinctions should be expressed linguistically at all. For example, in translating from Japanese to English, it is often necessary to supply determiners for noun phrases, and this in general cannot be done without deep understanding of the source text. Similarly, in translating from English to Japanese, politeness considerations, which in English are implicit in the social situation and explicit in very diffuse ways in, for example, the heavy use of hypotheticals, must be realized grammatically in Japanese. Machine translation therefore requires that the appropriate inferences be drawn and that the text be interpreted to some depth (see Oviatt, 1988). Recently, an elegant approach to inference in discourse interpretation has been developed at a number of sites (e.g., Hobbs et al., 1988; Charniak and Goldman, 1988; Norvig, 1987), all based on the notion of abduction, and we have begun to explore its potential application to machine translation. We argue that this approach provides the possibility of deep reasoning and of mapping between the languages at a variety of levels. (See also Kaplan et al., 1988, on the latter point.)[1]

Interpretation as Abduction. Abductive inference is inference to the best explanation. The easiest way to understand it is to compare it with

[1]The authors have profited from discussions about this work with Mark Stickel and with the participants in the Translation Group at CSLI. The research was funded by the Defense Advanced Research Projects Agency under Office of Naval Research contract N00014-85-C-0013, and by a gift from the Systems Development Foundation.

two words it rhymes with—deduction and induction. Deduction is when from a specific fact $p(A)$ and a general rule $(\forall x)p(x) \supset q(x)$ we conclude $q(A)$. Induction is when from a number of instances of $p(A)$ and $q(A)$ and perhaps other factors, we conclude $(\forall x)p(x) \supset q(x)$. Abduction is the third possibility. It is when from $q(A)$ and $(\forall x)p(x) \supset q(x)$, we conclude $p(A)$. Think of $q(A)$ as some observational evidence, of $(\forall x)p(x) \supset q(x)$ as a general law that could explain the occurrence of $q(A)$, and of $p(A)$ as the hidden, underlying specific cause of $q(A)$. Much of the way we interpret the world in general can be understood as a process of abduction.

When the observational evidence, the thing to be interpreted, is a natural language text, we must provide the best explanation of why the text would be true. In the TACITUS Project at SRI, we have developed a scheme for abductive inference that yields a significant simplification in the description of interpretation processes and a significant extension of the range of phenomena that can be captured. It has been implemented in the TACITUS System (Hobbs et al., 1988, 1990; Stickel, 1989) and has been applied to several varieties of text. The framework suggests the integrated treatment of syntax, semantics, and pragmatics described below. Our principal aim in this paper is to examine the utility of this framework as a model for translation.

In the abductive framework, what the interpretation of a sentence is can be described very concisely:

To interpret a sentence:

(1) Prove the logical form of the sentence,
    together with the constraints that
        predicates impose on their
        arguments,
    allowing for coercions,
Merging redundancies where possible,
Making assumptions where necessary.

By the first line we mean "prove from the predicate calculus axioms in the knowledge base, the logical form that has been produced by syntactic analysis and semantic translation of the sentence."

In a discourse situation, the speaker and hearer both have their sets of private beliefs, and there is a large overlapping set of mutual beliefs. An utterance stands with one foot in mutual belief and one foot in the speaker's private beliefs. It is a bid to extend the area of mutual belief to include some private beliefs of the speaker's. It is anchored referentially in mutual belief, and when we prove the logical form and the constraints, we are recognizing this referential anchor. This is the given information, the definite, the presupposed. Where it is necessary to make assumptions, the information comes from the speaker's private beliefs, and hence is the new information, the indefinite, the asserted. Merging redundancies is a way of getting a minimal, and hence a best, interpretation.

An Example. This characterization, elegant though it may be, would be of no interest if it did not lead to the solution of the discourse problems we need to have solved. A brief example will illustrate that it indeed does.

(2) The Tokyo office called.

This example illustrates three problems in "local pragmatics", the reference problem (What does "the Tokyo office" refer to?), the compound nominal interpretation problem (What is the implicit relation between Tokyo and the office?), and the metonymy problem (How can we coerce from the office to the person at the office who did the calling?).

Let us put these problems aside, and interpret the sentence according to characterization (1). The logical form is something like

(3) $$(\exists\, e, x, o, b) call'(e, x) \wedge person(x) \\ \wedge rel(x, o) \wedge office(o) \wedge nn(t, o) \\ \wedge Tokyo(t)$$

That is, there is a calling event e by a person x related somehow (possibly by identity) to the explicit subject of the sentence o, which is an office and bears some unspecified relation nn to t which is Tokyo.

Suppose our knowledge base consists of the following facts: We know that there is a person John who works for O which is an office in Tokyo T.

(4) $$person(J),\ work\text{-}for(J, O),\ office(O), \\ in(O, T),\ Tokyo(T)$$

Suppose we also know that $work\text{-}for$ is a possible coercion relation,

(5) $$(\forall\, x, y) work\text{-}for(x, y) \supset rel(x, y)$$

and that *in* is a possible implicit relation in compound nominals,

(6) $(\forall\, y,z)in(y,z) \supset nn(z,y)$

Then the proof of all but the first conjunct of (3) is straightforward. We thus assume $(\exists\, e)call'(e,J)$, and this constitutes the new information.

Notice now that all of our local pragmatics problems have been solved. "The Tokyo office" has been resolved to O. The implicit relation between Tokyo and the office has been determined to be the *in* relation. "The Tokyo office" has been coerced into "John, who works for the Tokyo office."

This is of course a simple example. More complex examples and arguments are given in Hobbs et al., (1990). A more detailed description of the method of abductive inference, particularly the system of weights and costs for choosing among possible interpretations, is given in that paper and in Stickel, (1989).

The Integrated Framework. The idea of interpretation as abduction can be combined with the older idea of parsing as deduction (Kowalski, 1980, pp. 52-53; Pereira and Warren, 1983). Consider a grammar written in Prolog style just big enough to handle sentence (2).

(7) $(\forall\, i,j,k)np(i,j) \wedge v(j,k) \supset s(i,k)$

(8) $(\forall\, i,j,k,l)det(i,j) \wedge n(j,k) \wedge n(k,l)$
$\supset np(i,l)$

That is, if we have a noun phrase from "inter-word point" i to point j and a verb from j to k, then we have a sentence from i to k, and similarly for rule (8).

We can integrate this with our abductive framework by moving the various pieces of expression (3) into these rules for syntax, as follows:

(9) $(\forall\, i,j,k,e,x,y,p)np(i,j,y) \wedge v(j,k,p)$
$\wedge\, p'(e,x) \wedge Req(p,x) \wedge rel(x,y)$
$\supset s(i,k,e)$

That is, if we have a noun phrase from i to j referring to y and a verb from j to k denoting predicate p, if there is an eventuality e which is the condition of p being true of some entity x (this corresponds to $call'(e,x)$ in (3)), if x satisfies the selectional requirement p imposes on its argument (this corresponds to $person(x)$), and if x is somehow related to, or coercible from, y, then there is an *interpretable* sentence from i to k describing eventuality e.

(10) $$(\forall i, j, k, l, w_1, w_2, z, y) det(i, j, the) \\ \wedge n(j, k, w_1) \wedge n(k, l, w_2) \wedge w_1(z) \wedge w_2(y) \\ \wedge nn(z, y) \supset np(i, l, y)$$

That is, if there is the determiner "the" from i to j, a noun from j to k denoting predicate w_1, and another noun from k to l denoting predicate w_2, if there is a z that w_1 is true of and a y that w_2 is true of, and if there is an nn relation between z and y, then there is an *interpretable* noun phrase from i to l denoting y.

These rules incorporate the syntax in the literals like $v(j, k, p)$, the pragmatics in the literals like $p'(e, x)$, and the compositional semantics in the way the pragmatics expressions are constructed out of the information provided by the syntactic expressions.

To parse with a grammar in the Prolog style, we prove $s(0, N)$ where N is the number of words in the sentence. To parse and interpret in the integrated framework, we prove $(\exists e)s(0, N, e)$.

An appeal of such declarative frameworks is their usability for generation as well as interpretation (Shieber, 1988). Axioms (9) and (10) can be used for generation as well. In generation, we are given an eventuality E, and we need to find a sentence with some number n of words that describes it. Thus, we need to prove $(\exists n)s(0, n, E)$. Whereas in interpretation it is the new information that is assumed, in generation it is the terminal nodes, like $v(j, k, p)$, that are assumed. Assuming them constitutes uttering them.

Translation is a matter of interpreting in the source language (say, English) and generating in the target language (say, Japanese). Thus, it can be characterized as proving for a sentence with N words the expression

(11) $$(\exists e, n) s_E(0, N, e) \wedge s_J(0, n, e)$$

where s_E is the root node of the English grammar and s_J is the root node of the Japanese.

Actually, this is not quite true. Missing in the logical form in (3) and in the grammar of (9) and (10) is the "relative mutual identifiability" relations that are encoded in the syntactic structure of sentences. For example, the office in (2) should be mutually identifiable once Tokyo is identified. In the absence of these conditions, the generation conjunct of (11) only says to express something true of e, not something that will enable the hearer to identify it. Nevertheless, the framework as it is developed so far will allow us to address some nontrivial problems in translation.

This point exhibits a general problem in translation, machine or human, namely, how literal a translation should be produced. We may think of this as a scale. At one pole is what our current formalization yields—a translation that merely says something true about the eventuality asserted in the source sentence. At the other pole is a translation that translates explicitly every property that is explicit in the source sentence. Our translation below of example (2) lies somewhere in between these two poles. Ideally, the translation should be one that will lead the hearer to the same underlying situation as an interpretation. It is not yet clear how this can be specified formally.

The Example Translated. An idiomatic translation of sentence (2) is

(12) Tokyo no office kara denwa ga ari-mashita.
Tokyo 's office from call Subj existed

Let us say the logical form is as follows:

(13) $aru(e) \wedge ga(d,e) \wedge denwa(d) \wedge kara(o,e)$
$\wedge office(o) \wedge no(t,o) \wedge Tokyo(t)$

A toy grammar plus pragmatics for Japanese, corresponding to the grammar of (9)-(10) is as follows[2]:

(14) $(\forall i,j,k,l,e,p) pp(i,j,e) \wedge pp(j,k,e)$
$\wedge v(k,l,p) \wedge p(e) \supset s(i,l,e)$

(15) $(\forall i,j,k,x,e,part) np(i,j,x)$
$\wedge particle(j,k,part) \wedge part(x,e)$
$\supset pp(i,k,e)$

(16) $(\forall i,j,k,l,x,y) np(i,j,y) \wedge particle(j,k,no)$
$\wedge np(k,l,x) \wedge no(y,x)$
$\supset np(i,l,x)$

(17) $(\forall i,j,w,x) n(i,j,w) \wedge w(x) \supset np(i,j,x)$

[2]For simplicity in this example, we are assuming the words of the sentences are given; in practice, this can be carried down to the level of characters.

$pp(i,j,e)$ means that there is a particle phrase from i to j with the missing argument e. $part$ is a particle and the predicate it encodes.

If we are going to translate between the two languages, we need axioms specifying the transfer relations. Let us suppose "denwa" is lexically ambiguous between the telephone instrument $denwa_1$ and the calling event $denwa_2$. This can be encoded in the two axioms

(18) $(\forall x) denwa_1(x) \supset denwa(x)$
(19) $(\forall x) denwa_2(x) \supset denwa(x)$

Lexical disambiguation occurs as a byproduct of interpretation in this framework, when the proof of the logical form uses one or the other of these axioms.

"Denwa ga aru" is an idiomatic way of expressing a calling event in Japanese. This can be expressed by the axiom

(20) $(\forall e,x) call'(e,x) \supset (\exists d) denwa_2(d)$
$\quad \wedge ga(d,e) \wedge aru(e)$

The agent of a calling event is also its source:

(21) $(\forall e,x) call'(e,x) \supset Source(x,e)$

We will need an axiom that coarsens the granularity of the source. If John is in Tokyo when he calls, then Tokyo as well as John is the source.

(22) $(\forall x,y,e) Source(x,e) \wedge in(x,y)$
$\quad \supset Source(y,e)$

If x works for y, then x is in y:

(23) $(\forall x,y) work\text{-}for(x,y) \supset in(x,y)$

Finally, we will need axioms specifying the equivalence of the particle "kara" with the deep case Source

(24) $(\forall y,e) Source(y,e) \equiv kara(y,e)$

and the equivalence between the particle "no" and the implicit relation in English compound nominals

(25) $(\forall x, y) nn(x, y) \equiv no(x, y)$

Note that these "transfer" axioms encode world knowledge (22 and 23), lexical ambiguities (18 and 19), direct relations between the two languages (20 and 25), and relations between the languages and deep "interlingual" predicates (21 and 24).

The proof of expression (11), using the English grammar of (9)-(10), the knowledge base of (4)-(6), the Japanese grammar and lexicon of (14)-(19), and the transfer axioms of (20)-(25), is shown in Figure 1. Boxes are drawn around the expressions that need to be assumed, namely, the new information in the interpretation and the occurrence of lexical items in the generation.

The axioms occur at a variety of levels, from the very superficial (axiom 25), to very language-pair specific transfer rules (axiom 20), to deep relations at the interlingual level (axioms 21-24). This approach thus permits mixing in one framework both transfer and interlingual approaches to translation. One can state transfer rules between two languages at various levels of linguistic abstraction, and between different levels of the respective languages. Such freedom in transfer is exactly what is needed for translation, especially for such typologically dissimilar languages as English and Japanese. It is thus possible to build a single system for translating among more than two languages in this framework, incorporating the labor savings of interlingual approaches while allowing the convenient specificities of transfer approaches.

We should note that other translations for sentence (2) are possible in different contexts. Two other possibilities are the following:

(26) Tokyo no office ga denwa shimashita.
Tokyo 's office Subj call did-Polite
The Tokyo office made [a/the] call.

(27) Tokyo no office kara no denwa ga arimashita.
Tokyo 's office from 's call Subj existed-Polite
There was the call from the Tokyo office (that we were expecting).

The difference between (12) and (26) is the speaker's viewpoint. The speaker takes the receiver's viewpoint in (12), while it is neutral between the caller and the receiver in (26). (27) is a more specific version of (12) where the call is mutually identifiable. All of (12), (26) and (27) are polite with the suffix "-masu". Non-polite variants are also possible translations.

On the other hand, in the following sentence

(28) Tokyo no office kara denwa shimashita.
Tokyo 's office from call did-Polite
[ ] made [a/the] call from the Tokyo office.

there is a strong inference that the caller is the speaker or someone else who is very salient in the current context.

The use of "shimashita" ("did") in (26) and (28) indicates the description from a neutral point of view of an event of some agent in the Tokyo office causing a telephone call to occur at the recipient's end. This neutral point of view is expressed in (26). In (28), the subject is omitted and hence must be salient, and consequently, the sentence is told from the caller's point of view. In (12) "ari-mashita" ("existed") is used, and since the telephone call exists primarily, or only, at the recipient's end, it is assumed the speaker, at least in point of view, is at the receiver's end.

Although we have not done it here, it looks as though these kinds of considerations can be formalized in our framework as well.

Hard Problems. If a new approach to machine translation is to be compelling, it must show promise of being able to handle some of the hard problems. We have identified four especially hard problems in translating between English and Japanese.

1. The lexical differences (that occur between any two languages).
2. Honorifics.
3. Definiteness and number.
4. The context-dependent "information structure".

The last of these includes the use of "wa" versus "ga", the order of noun phrases, and the omission of arguments.

These are the areas where one language's morphosyntax requires distinctions that are only implicit in the commonsense knowledge or context in the other language. Such problems cannot be handled by existing sentence-by-sentence translation systems without unnecessarily complicating the representations for each language.

In this short paper, we can only give the briefest indication of why we think our framework will be productive in investigating the first three of these problems.

Lexical Differences. Lexical differences, where they can be specified precisely, can be encoded axiomatically:

$$(\forall x) water(x) \wedge cool/cold(x) \equiv mizu(x)$$
$$(\forall x) water(x) \wedge warm/hot(x) \equiv yu(x)$$
$$(\forall x) watch(x) \equiv tokei(x) \wedge worn(x)$$
$$(\forall x) clock(x) \equiv tokei(x) \wedge \neg worn(x)$$

Information required for supplying Japanese numeral classifiers can be specified similarly. Thus the equivalence between the English "two trees" and the Japanese "ni hon no ki" can be captured by the axioms

$$(\forall x) tree(x) \supset cylindrical(x)$$
$$(\forall x) cylindrical(x) \equiv hon(x)$$

Honorifics. Politeness is expressed in very different ways in English and Japanese. In Japanese it is grammaticized and lexicalized in sometimes very elaborate ways in the form of honorifics. One might think that the problem of honorifics does not arise in most practical translation tasks, such as translating computer manuals. English lacks honorifics and in Japanese technical literature they are conventionalized. But if we are translating business letters, this aspect of language becomes very important. It is realized in English, but in a very different way. When one is writing to one's superiors, there is, for example, much more embedding of requests in hypotheticals. Consider for example the following English sentence and its most idiomatic translation:

Would it perhaps be possible for you to lend me your book?
Go-hon o kashite-itadak-e-masu ka.
Honorific-book Obj lending-receive-can- Polite ?

In Japanese, the object requested is preceded by the honorific particle "go", "itadak" is a verb used for a receiving by a lower status person from a higher status person, and "masu" is a politeness ending for verbs. In English, by contrast, the speaker embeds the request in various modals, "would", "perhaps", and "possible", and uses a more formal register than normal, in his choice, for example, of "perhaps" rather than "maybe".

The facts about the use of honorifics can be encoded axiomatically, with predicates such as $HigherStatus$, where this information is known. Since all knowledge in this framework is expressed uniformly in predicate calculus axioms, it is straightforward to combine information from different "knowledge sources", such as syntax and the speech act situation, into single rules. It is therefore relatively easy to write axioms that, for example, restrict the use of certain verbs, depending on the relative status of the agent and object, or the speaker and hearer. For example, "to give" is translated into the Japanese verb "kudasaru" if the giver is of higher status than the recipient, but into the verb "sashiageru" if the giver is of lower status. Similarly, the grammatical fact about the use of the suffix "-masu" and the fact about the speech act situation that speaker wishes to be polite may also be expressed in the same axiom.

We can also express the facts concerning the use of the honorific particle "o" (or "go") before nouns. There seem to be three classes of nouns in this respect. Some nouns, such as "cha" ("tea"), always take the particle ("o-cha"). Some nouns, especially loan words like "kōhī" ("coffee"), never take the particle. Other nouns, such as "bōshi" ("hat"), take the honorific prefix if the entity referred to belongs to someone of higher status. For this class of nouns we can state the condition formally.

$$\begin{aligned}
&(\forall\, i, j, k, p, x, y) Honorific(i, j, o) \\
&\qquad \wedge\, Noun(j, k, p) \;\wedge\; p(x) \;\wedge\; possess(y, x) \\
&\qquad \wedge\, HigherStatus(y, Speaker) \\
&\quad \supset\; NP(i, k, x)
\end{aligned}$$

That is, if the honorific particle "o" occurs from point i to point j, the noun denoting the predicate p occurs from point j to point k, and p is true of some entity x where someone y possesses x and y is of higher status than the speaker, then there is an interpretable noun phrase from point i to point k referring to x.

Definiteness and Number. The definiteness and number problem is illustrated by the fact that the Japanese word "ki" can be translated into "the tree" or "a tree" or "the trees" or "trees". It is not so straightforward to deal with this problem axiomatically. Nevertheless, our framework, based as it is on deep interpretation and on the distinction between given and new information, provides us with what we need to begin to address the problem. A first approximation of a method for translating Japanese NPs into English NPs is as follows:

1. Resolve deep, i.e., find the referrent of the Japanese NP.

2. Does the Japanese NP refer to a set of two or more? If so, translate it as a plural, otherwise as a singular.

3. Is the entity (or set) "mutually identifiable"? If so, then translate it as a definite, otherwise as an indefinite.

"Mutually identifiable" means first of all that the description provided by the Japanese NP is mutually known, and secondly that there is a single most salient such entity. "Most salient" means that there are no other equally high-ranking interpretations of the Japanese sentence that resolve the NP in some other way. (Generic definite noun phrases are beyond the scope of this paper.)

Conclusion. We have sketched our solutions to the various problems in translation with a fairly broad brush in this short paper. We recognize that many details need to be worked out, and that in fact most of the work in machine translation is in working out the details. But we felt that in proposing a new formalism for translation research, it was important to stand back and get a view of the forest before moving in to examine the individual trees.

Most machine translation systems today map the source language text into a logical form that is fairly close to the source language text, transform it into a logical form that is fairly close to a target language text, and generate the target language text. What is needed is first of all the possibility of doing deep interpretation when that is what is called for, and secondly the possibility of translating from the source to the target language at a variety of levels, from the most superficial to levels requiring deep interpretation and access to knowledge about the world, the context, and the speech act situation. This is precisely what the framework we have presented here makes possible.

References

[1] Charniak, Eugene, and Robert Goldman, 1988. "A Logic for Semantic Interpretation", *Proceedings, 26th Annual Meeting of the Association for Computational Linguistics*, pp. 87-94, Buffalo, New York, June 1988.

[2] Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards, 1988. "Interpretation as Abduction", *Proceedings, 26th Annual Meeting of the*

Association for Computational Linguistics, pp. 95-103, Buffalo, New York, June 1988.

[3] Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards, 1990. "Interpretation as Abduction", forthcoming technical note.

[4] Kaplan, R. M., K. Netter, J. Wedekind, A. Zaenen, 1989. "Translation by Structural Correspondences", *Proceedings*, Fourth Conference of the European Chapter of the Association for Computational Linguistics, Manchester, United Kingdom, pp. 272-281.

[5] Kowalski, Robert, 1980. *The Logic of Problem Solving*, North Holland, New York.

[6] Norvig, Peter, 1987. "Inference in Text Understanding", *Proceedings, AAAI-87, Sixth National Conference on Artificial Intelligence*, Seattle, Washington, July 1987.

[7] Oviatt, Sharon L., 1988. "Management of Miscommunications: Toward a System for Automatic Telephone Interpretation of Japanese-English Dialogues", Technical Note 438, SRI International, May 1988.

[8] Pereira, Fernando C. N., and David H. D. Warren, 1983. "Parsing as Deduction", *Proceedings of the 21st Annual Meeting, Association for Computational Linguistics*, pp. 137-144. Cambridge, Massachusetts, June 1983.

[9] Shieber, Stuart M., 1988. "A Uniform Architecture for Parsing and Generation", *Proceedings, 12th International Conference on Computational Linguistics*, pp. 614-619, Budapest, Hungary.

[10] Stickel, Mark E. 1989. "A Prolog Technology Theorem Prover: A New Exposition and Implementation in Prolog", Technical Note No. 464. Menlo Park, Calif.: SRI International.

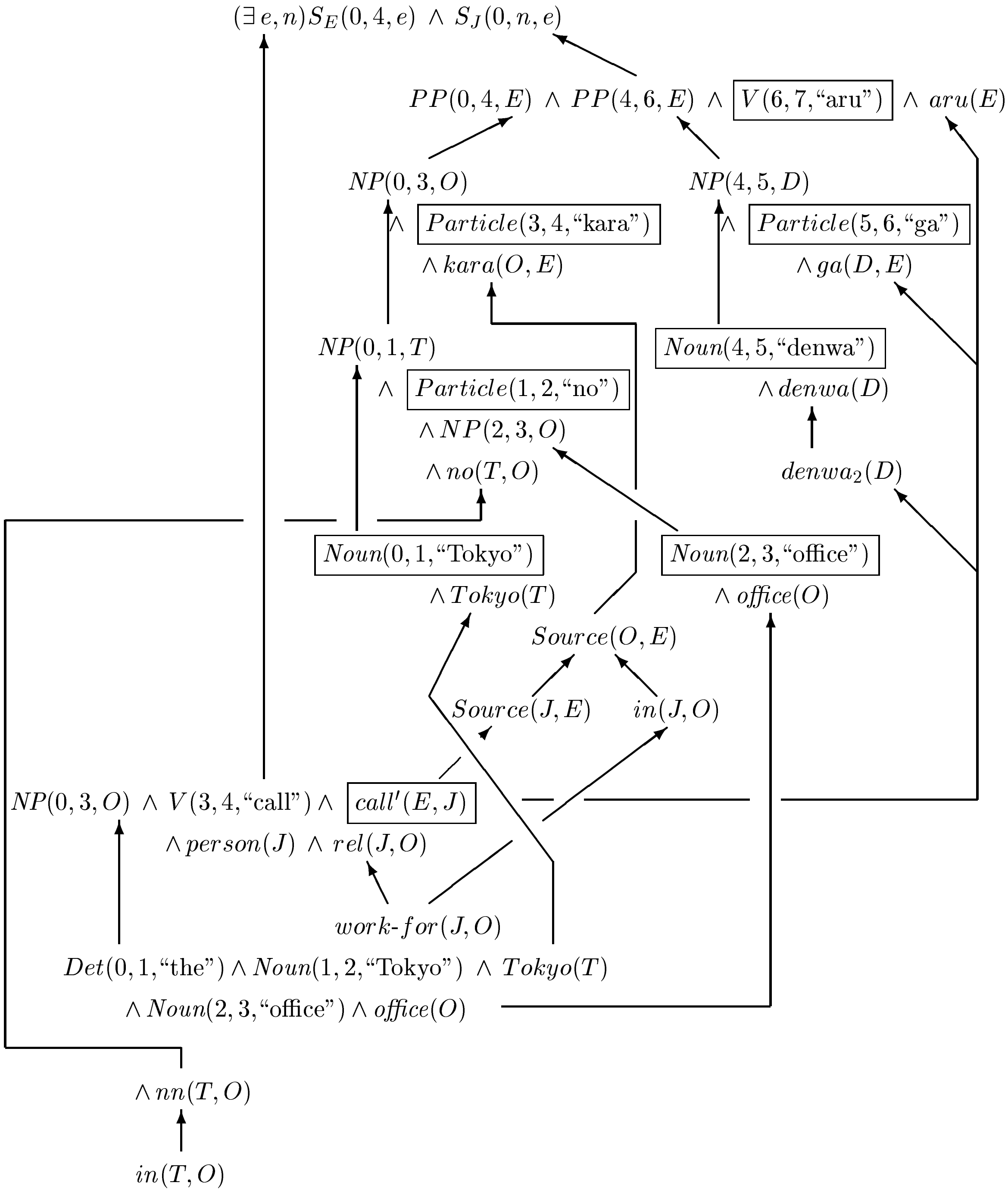

Figure 1: Translation by Abduction