Defining the Semantics of Verbal Modifiers in the Domain of Cooking Tasks

Robin F. Karlin
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389

Abstract

SEAFACT (Semantic Analysis For the Animation of Cooking Tasks) is a natural language interface to a computer-generated animation system operating in the domain of cooking tasks. SEAFACT allows the user to specify cooking tasks using a small subset of English. The system analyzes English input and produces a representation of the task which can drive motion synthesis procedures. This paper describes the semantic analysis of verbal modifiers on which the SEAFACT implementation is based.

Introduction

SEAFACT is a natural language interface to a computer-generated animation system (Karlin, 1988). SEAFACT operates in the domain of cooking tasks. The domain is limited to a mini-world consisting of a small set of verbs chosen because they involve rather complex arm movements which will be interesting to animate. SEAFACT allows the user to specify tasks in this domain, using a small subset of English. The system then analyzes the English input and produces a representation of the task. An intelligent simulation system (Fishwick, 1985,1987), which is currently being extended, will provide the final link between the SEAFACT representation and lower level motion synthesis procedures. The representation consists of a decomposition of verbs into primitive actions which are semantically interpretable by the motion synthesis procedures. It also includes default information for all knowledge which is not made explicit in the input, but must be explicit in the animated output. The representation contains sufficient non-geometric information needed to schedule task start and end times, describe concurrent actions, and provide reach, grasp, and motion goals.

An empirical, linguistic study of recipes was conducted with the goals of delimiting the scope of the cooking domain, identifying important verbal modifiers, and defining the semantics of those modifiers. This paper is concerned primarily with describing the results of this study and the implementation of some of the modifiers.

A Linguistic Analysis of Verbal Modifiers

An empirical study of approximately 110 sentences from nine cookbooks was carried out. Verbal modifiers were found to play an essential role in the expressive power of these sentences. Therefore, in order to develop a representation for the verbal modifiers, the study describes and categorizes their occurences and provides a semantic analysis of each of the categories. Each of the categories is considered a semantic role in the representation of the natural language input. Temporal adverbials were found to be particularly prevalent in recipes because they are needed to specify temporal information about actions which is not inherent in the meaning of verbs and their objects. This paper discusses two categories of temporal modifiers: duration and repetitions as well as speed modifiers. Other categories of modifiers which were analyzed include quantity of the object, end result, instrument, and force.

Passonneau (1986) and Waltz (1981,1982) are concerned with developing semantic representations adequate for representing adverbial modification. Passonneau's work shows that to account for tense and grammatical aspect requires a much more complex representation of the temporal components of language than the one used in SEAFACT. However, she does not look at as many categories of temporal adverbials, nor does she propose a specific representation for them. Waltz (1982) suggests that adverbs will be represented by the scales in his event shape diagrams. For example, time adverbials will be rep-

resented by the time scale and quantity adverbials by the scale for quantity of the verbal objects. This is similar to the approach taken in SEAFACT. In SEAFACT scales are replaced by default amounts for the category in question, for example the duration of a primitive action.

Aspectual Category of an Event

The aspectual category of an event is relevant because it affects which types of modifiers (e.g., repetitions, duration) can co-occur with the event. The analysis of aspect given in Moens (1987) (see also (Moens, 1988)) is adopted here. Moens and Steedman identify temporal/aspectual types following Vendler, but introduce new terminology. They apply these types to entire sentences, analyzed in their global context. Moens and Steedman's events are classified as culminated processes, culminations, points, or processes. The majority of events in the cooking domain are *culminated processes*. A culminated process is

> a state of affairs that also extends in time but that *does* have a particular culmination associated with it at which a change of state takes place. (Moens, 1987, p. 1)

Each process in cooking must have a culmination because any cooking task involves a finite sequence of steps, whose goal is to bring about a state change. An important point about verbal modifiers in the cooking domain, revealed in this study, is that many of them are concerned with characterizing the culmination points of processes. In many cases a verb and object alone do not specify a clear culmination point. For example, the command **beat the cream** does not contain information about the desired culmination of the process, that is, when to stop the beating. Some sort of verbal modifier such as **for 10 minutes** or **just until it forms peaks** is necessary to specify the culmination of the process.

Another aspectual type is a *culmination*. A culmination is

> an event which the speaker views as accompanied by a transition to a new state of the world. This new state we will refer to as the "consequent state" of the event. (Moens, 1987, p. 1)

Culminations, such as **cover the pot**, are not extended in time as are processes and culminated processes.

In addition to the sentential aspect discussed above, the SEAFACT implementation identifies the lexical aspect of the verb. The lexical aspect refers to the aspectual category which can be ascribed to a verb considered outside of an utterance. For example, the lexical aspect of the verb *stir* is a process. However, the sentential aspect of the sentence *stir the soup for 3 minutes* is a culminated process. The implementation checks that the sentential aspect of each input sentence containing a process verb is a culminated process. That is, there must be some verbal modifier which coerces the process into a culminated process. If this is not the case, as in the sentence *stir the soup*, then the input is rejected since it would specify an animation without an ending time. The lexical aspect is also used in the analysis of speed modifiers, as discussed below.

The Number of Repetitions of the Action

Any expression which includes an endpoint, and therefore belongs to one of the aspectual classes of points, culminations, or culminated processes can be described as having a number of discrete repetitions. When a culminated process is described as having a number of repetitions, it is the entire process which is repeated. Process type events cannot have a number of repetitions associated with them since they do not include the notion of an end point. The number of repetitions of the event can be specified as a cardinal number, as a frequency, or indirectly as a result of the object of the verb being plural, having multiple parts, or being a mass term.

Cardinal Count Adverbials

Cardinal count adverbials (Mourelatos, 1981, p. 205) specify an exact number of repetitions of the event.

(1) **baste** *twice* **during the cooking period** (Rombauer, 1931, p. 350)

Notice that in the case of certain verbs or sentential contexts it is not possible to specify a number of repetitions for a culminated process. This is the case when the culmination involves a state change to the object which makes a repetition of the action impossible or meaningless. Consider the example, *Freeze twice. Freeze is a culminated process and once the culmination has taken place the new state of the substance makes a repetition of the process redundant. Talmy (1985) proposes a classification scheme of aspectual types of verb roots which formalizes this distinction. He would classify freeze as a *one-way non-resettable* verb and baste as a *one-way resettable verb* (Talmy, 1985, p. 77) He suggests that these types can be distinguished by their ability to appear with iterative

expressions. This distinction can also be made by means of world knowledge about the verbs in question.

Frequency Adverbials

Frequency adverbials (Mourelatos, 1981, p. 205) describe the number of repetitions of an action using a continuous scale with gradable terms (Croft, 1984, p. 26) such as frequently, occasionally, and seldom.

(2) **Bring to a boil, reduce the heat, and simmer 20 minutes, stirring** *occasionally,* **until very thick.** (Poses, 1985, p. 188)

The meaning of frequency adverbials is best captured by stating the length of the intervals between repetitions of the action. For example, the meaning of occasionally is that the number of minutes between incidents of stirring is large. An additional complication is that frequency adverbials must be interpreted relative to the total length of time during which the event may be repeated. If the total time period is longer, the intervals must be proportionately longer.

Like other gradable terms, such as tall and short, frequency adverbials are interpreted relative to their global context, in this case the cooking domain. Values must be determined for each of the gradable terms, based on knowledge of typical values in the domain. In the SEAFACT implementation these values consist of cardinal numbers which specify the length of an interval between repetitions of the action, expressed as a percentage of the total time period.

The following calculations are made when a frequency adverbial is present in a sentence. The length of a single interval between incidents of the action is calculated by using a percentage value associated with the frequency adverbial, such that IntervalTime = Percentage X TotalTime. The number of intervals present during the total time period is calculated by dividing the total time period by the sum of the length of one incident of the action and the length of a single interval.

A simplifying assumption is made here that the intervals between repetitions are equal. Occasionally might then mean intervals which are 25 per cent of the total time period, and frequently might mean intervals which are 5 per cent of the total time period. This algorithm seems to coincide with the intuitive judgment that it is not normal to say stir occasionally during a very short time period such as 30 seconds. In such a case, the length of an individual stirring event might be longer than the total time. That is, for the domain in question there is some minimum interval between stirring events which is necessary for the term occasionally to be appropriate.

Plural Objects

The use of plural objects or mass terms with a verb may or may not indicate that the action is to be repeated. The verb may indicate a single action which is performed on multiple objects simultaneously, or it may indicate an action which is repeated for each of a number of objects. This distinction does not always coincide with a mental conception of the objects as a mass or as individuals. Rather, it depends on physical attributes of the objects such as size and consistency.

(3) **chop the nuts**

In (3), world knowledge tells us that since nuts are small and relatively soft they can be chopped together in a group, perhaps using a cleaver.

(4) **chop the tomatoes with a knife**

Here, world knowledge tells us that (4) usually requires a separate chopping event for each tomato, since tomatoes are large compared to knives and have skins which are not easily pierced. Notice that this is a case of repetition of a culminated process. Verbal modifiers may also be used to make explicit whether an action is to be performed separately on each object in a group or once on a group of objects together.

(5) **beat in the eggs** *one at a time* (Gourmet, 1986, p. 12)

(6) **beat in 5 eggs until smooth**

In (5), the phrase one at a time makes explicit that there is to be a separate beating process for each egg. In (6), a sentence without a verbal modifier, the culminated process beat in is performed once on the objects indicated.

The Duration of an Action

Any expression whose aspectual type is a process or culminated process can co-occur with a duration modifier. The duration of a culminated process refers to the amount of time it continues before the culmination of the process. Duration can be specified as a cardinal number or a gradable term, corresponding to the categories used for number of repetitions. Duration can also be specified as co-extensive with the duration of another event, in terms of the change which signals the culmination, and as a disjunction of an explicit duration and a state change.

Explicit Duration in Time Units

Verbal modifiers may specify an explicit duration by giving a length of time. This can be less exact when a range of time or a minimum is specified.

(7) stir *for 1 minute;* set aside. (Morash, 1982, p. 132)

Duration Given by Gradable Terms

The duration of an action can be specified by gradable terms on a continuous scale.

(8) blend *very briefly* (Robertson, 1976, p. 316)

Duration Co-extensive with the Duration of Another Action

In the cooking domain it is often necessary to do several actions simultaneously. In such cases it is most natural to express the duration of one of the activities in terms of the duration of the other one.

(9) Continue to cook *while gently folding in the cheeses* with a spatula. (Poses, 1985, p. 186)

(10) Reduce the heat to medium and fry the millet, *stirring,* for 5 minutes or until it is light golden. (Sahni, 1985, p. 283)

Duration Characterized by a State Change

All processes in the cooking domain must have culminations since cooking consists of a finite number of steps executed with limited resources. The language used to describe these processes can convey their culminations in different ways. In some cases a verb may contain inherent information about the endpoint of the action which it describes. In other cases verbal modifiers characterize the endpoint.

(11) Chop the onion.

Example (11) specifies a culminated process whose endpoint is defined by the state of the onion. While the desired final state of the onion could be specified more exactly by some adverb such as finely or coarsely, in the absence of such a modifier an endpoint can be established based on lexical knowledge about the state of an object which has been chopped.

In many cases, however, the meaning of the process verb does not include information on the endpoint of the process, or the domain requires more specific information than that conveyed by the verb alone. For example, in many contexts, the verb beat does not supply the duration or the particular end result of the beating which would determine the duration. This is because different amounts of beating bring about different final states for many substances.

Therefore, the cooking domain includes many examples of duration of an action characterized by the

specification of a state change in the object being acted on. There must be some perceptual test which verifies when a state change has occurred. For visual changes the test consists of looking at the substance in question. A preparatory action is required only if the substance is not immediately visible, for example, if it is in the oven or in a closed pot. Changes which must be perceived by other senses, usually require additional actions. For example, to perform a tactile test one must touch the substance either directly or with some instrument.

The following is an example of a state change which can be perceived visually without an active test.

(12) Saute over high heat *until moisture is evaporated* (Morash, 1982, p. 131)

Disjunctions of Explicit Durations and State Changes

(13) steam *2 minutes or until mussels open* (Poses, 1985, p. 83)

The meaning of sentences in this category is not the same as that of logical disjunction. Example (13) does not give the cook a choice between steaming for 2 minutes or until the mussels open. The actual meaning of these disjunctions is that the state change is to be used to determine the duration of the action. The explicit duration provides information on the usual amount of time that is needed for the state change to take place.

Ball (1985) discusses problems that arise in the semantic interpretation of what she calls metalinguistic or non-truth functional disjunction. "The first clause is asserted, and the right disjunct provides an alternate, more accessible description of the referent of the left disjunct." (Ball, 1985, p. 3) The truth of these sentences depends on the truth of the first disjunct. Ball claims that if the first disjunct is true and the second is not, then the sentence is still true although "our impression will be that something has gone wrong." (Ball, 1985, p. 3)

The disjunctions of explicit durations and state changes seem to be another type of metalinguistic disjunction. They are very similar to the examples given by Ball except that it is the right disjunct which determines the truth of the sentence and the left disjunct which provides an alternate description. Furthermore, this alternate does not have to be strictly synonymous with the right disjunct. The semantics of these disjunctions includes the notion that the left disjunct is only an approximation.

The Speed

The following verbal modifiers are gradable terms which characterize the speed of the action.

(14) *quickly* tilt and turn the dish (Heatter, 1965, p. 400)

(15) *very gradually* pour (Heatter, 1965, p. 393)

The SEAFACT implementation contains values for these terms based on knowledge of typical values in the domain. These values are the amount by which the default duration of an action should be multiplied to arrive at the new duration specified by the speed term.

The lexical aspect of the verb is used to decide whether all or only a portion of the primitive actions which comprise the verbal action are affected by the speed factor. If the verb is a process then only a portion of the primitive actions are affected. For example, *stir the soup quickly for 5 minutes* means to make the repeated rotations of the instrument quickly, probably in order to prevent the soup from burning. It does not imply that the entire motion associated with stirring, which includes picking up the instrument and putting it in the soup and later removing it from the soup, must be done quickly. The latter interpretation would mean that the speed term was meant to modify the time which the entire action takes to complete. However, processes in this domain must be specified with a duration and so the duration of the entire action is already fixed.

In contrast, if the lexical aspect of the verb is a culmination or culminated process then the duration of the entire action is meant to be modified by the speed term. An example of this is *cover the pot quickly.*

The SEAFACT Implementation

There are several stages in the translation from English input to the final representation required by the animation simulator. The first stage includes parsing and the production of an intermediate semantic analysis of the input. This is accomplished by BUP, A Bottom Up Parser (Finin, 1984). BUP accepts an extended phrase structure grammar. The rules consist of the intermediate semantic representation and tests for rule application. The latter include selectional restrictions which access information stored in several knowledge bases. The intermediate semantic representation consists of roles and their values, which are taken from the input sentence.

SEAFACT includes a number of knowledge bases which are implemented using DC-RL, a frame-based knowledge representation language (Cebula, 1986). Two of these knowledge bases, the Object KB and the Linguistic Term KB, are used by the parser to enforce selectional restrictions attached to the grammatical rules.

The Object KB contains world knowledge about the objects in the domain. It contains a representation of each object which can be referred to in the natural language input. These objects are classified according to a very general conceptual structure. For example, all edible items are classified as food, cooking tools are classified as instruments, and cooking vessels are classified as containers. This information is used to enforce selectional restrictions in the rules for prepositional phrases. The selectional restrictions check the category to which the prepositional object belongs. For example, if the prepositional object is an instrument then the rule which applies builds an intermediate semantic representation of the form (INSTRUMENT *prepositional-object*). If the prepositional object denotes a time, and the preposition is for, then the rule which applies builds an intermediate semantic representation of the form (DURATION (EXPLICIT *prepositional-object*)).

The Linguistic Term KB contains a classification of adverbial modifiers which is used to enforce selectional restrictions on the rules for adverbial phrases. For example, if an adverb is classified as a *frequency term* then the rule which applies builds an intermediate semantic representation of the form (REPETITIONS (FREQUENCY *frequency-term*)).

The second stage in the processing is to create representations for the verb and the event. The event representation has roles for each of the temporal verbal modifiers. Each verb has its own representation containing roles for each of the verbal modifiers which can occur with that verb. The verb representations contain default values for any roles which are *essential* (Palmer, 1985). Essential roles are those which must be filled but not necessarily from the input sentence. For example, the representation for the verb **stir** includes the essential role **instrument** with a default value of **spoon**. After the event and verb representations are created, the role values in those representations are filled in from the roles in the intermediate semantic representation. Default values are used for any roles which were not present in the input sentence.

Each verb in the input is represented by a number of primitive actions which are interpretable by the animation software. In the second stage, the system also creates a representation of the final output which includes values for the starting time and duration of each of these actions.

The third stage in the processing is accomplished by the Modifier Analysis Component (MAC). This function performs the additional processing required by some of the temporal verbal modifiers such as frequency terms. This processing consists of modifying the output to reflect the temporal modifiers. This may mean changing the duration of actions (for speed and duration modifiers), modifying the number of times the output is repeated (for repetition modifiers), or interspersing intervals of no action with the intervals of action (for frequency modifiers).

The final output is created by filling in the primitive action representations with values from the verb and event representations.

Consider how SEAFACT processes two example sentences. In the first example, *Stir the batter with a wisk for 2 minutes*, the intermediate semantic representation includes a substance1 role filled by *batter*, an instrument role filled by *wisk*, and a duration role filled by *2 minutes*. These values are inserted in the verb and event representations for the sentence. The MAC modifies the duration of the primitive actions which make up stir so that the duration of the total stirring event is 2 minutes.

The second example, *Stir the soup occasionally for 2 minutes* is more complicated because of the frequency adverbial. The intermediate semantic representation includes a substance1 role filled by *soup*, a duration role filled by *2 minutes*, and a repetitions role filled by *occasionally*. These values are inserted in the verb and event representations. The default value for the instrument role, *spoon*, is used. The MAC finds the frequency adverbial and checks for the presence of a duration. However, if no duration were specified, then the sentence would be rejected because the animation requires that each action be finite. The duration specifies the total time interval during which the frequency adverbial applies. The algorithm described above is used to compute the length of the intervals between stirring events. The length of a single stirring event is a default which is part of the representation of the primitive actions. The number of stirring events which fit in the total time period is calculated. The output consists of repetitions of pairs of the following type: the primitives for a stirring event and a specification for no action during the interval between stirring events. A planner could be used to insert some other action into the intervals of no action.

Conclusion

This analysis has identified categories of verbal modifiers which are found frequently in recipes. While all of these categories are found in other domains as well, some of them are particularly prevalent in this domain because the purpose of recipes is to describe procedures. The temporal category which characterizes the duration of an action by a state change is particularly common in recipes for two reasons. First, the physical process of cooking always involves state changes to objects and second, the meaning of many verbs used to describe cooking processes does not include information about the state change which should trigger the culmination of the process. Therefore, verbal modifiers are necessary to make the desired state changes explicit.

This analysis has also shown a relationship between aspectual categories of events and the modifiers which may co-occur with them. For example, the categories of modifiers which express the number of repetitions of an action can only modify expressions which include an endpoint, that is, points, culminations, or culminated processes.

The analysis of the verbal modifier categories reveals many areas where common sense knowledge or physical knowledge about the world is required to represent the semantics of these categories. For example, when an action is performed on a plural object, physical knowledge about the size and consistency of the objects and about the action itself is necessary to tell us whether it must be repeated for each of the objects separately or performed on all the objects in a group.

SEAFACT is a successful implementation of a natural language interface to a computer-generated animation system, operating in the domain of cooking tasks. The primitive actions along with the timing information in the SEAFACT output are used to represent the range of verbal modifiers discussed in this paper. The output will be interpreted by an interface to the lower level motion synthesis procedures. This interface (Badler, 1988, 1987a, 1987b) can interpret each type of information in the SEAFACT output: motion changes (e.g. rotation), motion goals, constraints in position and orientation, and temporals.

Acknowledgements

66

References

Badler, Norman I., Jeffrey Esakov, Diana Dadamo, and Phil Lee, Animation Using Constraints, Dynamics, and Kinematics, in preparation, Technical Report, Department of Computer and Information Science, University of Pennsylvania, 1988.

Badler, Norman I., Computer Animation Techniques, in *2nd International Gesellschaft für Informatik Congress on Knowledge-Based Systems*, Springer-Verlag, Munich, Germany, October 1987a, pp. 22-34.

Badler, Norman I., Kamran Manoochehri, and Graham Walters, Articulated Figure Positioning by Multiple Constraints, *IEEE Computer Graphics and Applications*, June 1987b, pp. 28-38.

Ball, Catherine N., On the Interpretation of Descriptive and Metalinguistic Disjunction, unpublished paper, University of Pennsylvania, August 1985.

Cebula, David P., The Semantic Data Model and Large Data Requirements, University of Pennsylvania, CIS Dept.,Technical Report 87-79, Sept 1986.

Croft, William, The Representation of Adverbs, Adjectives and Events in Logical Form, Technical Note 344, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, Menlo Park, Ca, December 1984.

Finin, Tim and Bonnie Lynn Webber, BUP A Bottom Up Parser, Technical Report MC-CIS-83-16, University of Pennsylvania, 1984.

Fishwick, Paul A., The Role of Process Abstraction in Simulation, submitted to *IEEE Systems, Man and Cybernetics*, April 1987.

Fishwick, Paul A., Hierarchical Reasoning: Simulating Complex Processes Over Multiple Levels of Abstraction, PhD Thesis, Technical Report, University of Pennsylvania, 1985.

Gourmet Magazine, Volume XLVI, Number 6, June 1986.

Karlin, Robin F., SEAFACT: Semantic Analysis for the Animation of Cooking Tasks, Technical Report, MS-CIS-88-04, Graphics Lab 19, Computer and Information Science, University of Pennsylvania, 1988.

Moens, Marc and Mark Steedman, Temporal Ontology in Natural Language, in *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, ACL, 1987, pp. 1-7.

Moens, Marc and Mark Steedman, forthcoming, *Computational Linguistics*, Volume 14, Number 2, 1988.

Morash, Marion *Victory Garden Cookbook*, Alfred A. Knopf, N.Y., 1982.

Mourelatos, Alexander P. D., Events, Processes, and States, in *Syntax and Semantics, Tense and Aspect*, Vol. 14, Philip Tedeschi and Annie Zaenen (eds.), Academic Press, New York, 1981, pp. 191-212.

Palmer, Martha S., Driving Semantics for a Limited Domain, PHD Dissertation, University of Edinburgh, 1985.

Passonneau, Rebecca J., A Computational Model of the Semantics of Tense and Aspect, forthcoming, *Computational Linguistics*, Volume 14, Number 2, 1988, Tech. Memo 43, Dec. 17, 1986, Unisys, Paoli Research Center, Paoli, Pa, Dec. 1986.

Poses, Steven, Anne Clark, and Becky Roller, *The Frog Commissary Cookbook*, Doubleday & Company, Garden City, N.Y., 1985.

Rombauer, Irma S. and Marion Rombauer Becker, *Joy of Cooking*, Signet, New American Library, N.Y., 1931.

Sahni, Julie, *Classic Indian Vegetarian and Grain Cooking*, William Morrow and Co., Inc., N.Y., 1985.

Talmy, Leonard, Lexicalization Patterns: Semantic Structure in Lexical Forms, in *Language typology and syntactic description, Volume III, Grammatical categories and the lexicon*, Timothy Shopen (ed.), Cambridge University Press, Cambridge, 1985.

Waltz, David L., Event Shape Diagrams, in *AAAI-1982*, pp. 84-87.

Waltz, David L., Toward a Detailed Model of Processing For Language Describing the Physical World, in *IJCAI-1981*, pp. 1-6.