A CONNECTIONIST MODEL OF SOME ASPECTS OF ANAPHOR RESOLUTION

Ronan G. Reilly
Educational Research Centre
St Patrick's College, Drumcondra
Dublin 9, Ireland

ABSTRACT

This paper describes some recent developments in language processing involving computational models which more closely resemble the brain in both structure and function. These models employ a large number of interconnected parallel computational units which communicate via weighted levels of excitation and inhibition. A specific model is described which uses this approach to process some fragments of connected discourse.

I CONNECTIONIST MODELS

The human brain consists of about 100,000 million neuronal units with between a 1000 and 10,000 connections each. The two main classes of cells in the cortex are the striate and pyramidal cells. The pyramidal cells are generally large and heavily arborized. They are the main output cells of a region of cortex, and they mediate connections between one region and the next. The striate cells are smaller, and act more locally. The neural circuitry of the cortex is, apart from some minor variations, remarkably consistent. Its dominant characteristics are its parallelism, its large number processing units, and the extensive interconnection of these units. This is a fundamentally different structure from the traditional von Neumann model. Those in favor of adopting a connectionist approach to modelling human cognition argue that the structure of the human nervous system is so different from the structure implicit in current information-processing models that the standard approach cannot ultimately be successful. They argue that even at an abstract level, removed from immediate neural considerations, the fundamental structure of the human nervous system has a pervasive effect.

Connectionist models form a class of spreading activation or active semantic network model. Each primitive computing unit in the network can be thought of as a stylized neuron. Its output is a function of a vector of inputs from neighbouring units and a current level of excitation. The inputs can be both excitatory and inhibtory. The output of each unit has a restricted range (in the case of the model described here, it can have a value between 1 and 10). Associated with each unit are a number of computational functions. At each input site there are functions which determine how the inputs are to be summarized. A potential function determines the relationship between the summarized site inputs and the unit's overall potential. Finally, an output function determines the relationship between a unit's potential and the value that it transmits to its neighbours.

There are a number of constraints inhererent in a neurally based model. One of the most significant is that the coinage of the brain is frequency of firing. This means that the inputs and outputs cannot carry more than a few bits of information. There are not enough bits in firing frequency to allow symbol passing between individual units. This is perhaps the single biggest difference between this approach and and that of standard information-processing models. Another important constraint is that decisions in the network are completely distributed, each unit computes its output solely on the basis of its inputs; it cannot "look around" to see what others are doing, and no central controller gives it instructions.

A number of language related applications have been developed using this type of approach. The most notable of these is the model of McClelland and Rumelhart (1981). They demonstrated that a model based on connectionist principles could reproduce many of the characteristcs of the so-called word-superiority effect. This is an effect in which letters in briefly presented words and pseudo-words are more easily identifiable than letters in non-words. At a higher level in the processing hierarchy, connectionist schemes have been proposed for modelling word sense disambiguation (Cottrell & Small, 1983), and for sentence parsing in general (Small, Cottrell, & Shastri, 1982).

The model described in this paper is basically an extension of the work of Cottrell and Small (1983), and of Small (1982). It extends their sentence-centred model to deal with connected text, or discourse, and specifically with anaphoric resolution in discourse. The model is not proposed as definitive in any way. It merely sets out to illustrate the properties of connectionist models, and to show how such models might be extended beyond simple word recognition applications.

II ANAPHORA

The term anaphor derives from the Greek for "pointing back". What is pointed to is often referred to as the antecedent of the anaphor. However, the precise definition of an antecedent is problematic. Superficially, it might be thought of as a preceding text element. However, as Sidner (1983) pointed out words do not refer to other words; people use words to refer to objects, and anaphora are used to refer to objects which have already been mentioned in a discourse. Sidner also maintains that the concept of co-reference is inadequate to explain the relationship between anaphor and antecedent. Co-reference means that anaphor and antecedent both refer to the same object. This explanation suffices for a sentence like:

(1) I think green apples are best and they make the best cooking apples too.

where both they and green apples refer to the same object. However, it is inadequate when dealing with the following discourse:

(2) My neighbour has an Irish Wolfhound. They are really huge, but friendly dogs.

In this case they refers to the class of Irish Wolfhounds, but the antecedent phrase refers to a member of that set. Therefore, the anaphor and antecedent cannot be said to co-refer. Sidner introduces the concept of specification and co-specification to get around this problem. Instead of referring to objects in the real world, the anaphor and its antecedent specify a cognitive element in the hearer's mind. Even though the same element is not co-specified one specification may be used generate the other. This is not possible with co-reference because, as Sidner puts it:

Co-specification, unlike co-reference, allows one to construct abstract representations and define relationships between them which can be studied in a computational framework. With coreference, no such use is possible, since the object referred to exists in the world and is not available for examination by the computational process. (Sidner, 1983; p. 269).

Sidner proposes two major sources of constraint on what can become the co-specification of an anaphoric reference. One is the shared knowledge of speaker and hearer, and the other is the concept of focus. At any given time the focus of a discourse is that discourse element which is currently being elaborated upon, and on which the speakers have centered their attention. This concept of focus will be implemented in the model to be described, though differently from the way Sidner (1983) has envisaged it. In her model possible focuses are examined serially, and a decision is not made until a sentence has been completely analyzed. In the model proposed here, the focus is arrived at on-line, and the process used is a parallel one.

III THE SIMULATOR

The model described here was constructed using an interactive connectionist simulator written in Salford LISP and based on the design for the University of Rochester's ISCON simulator (Small, Shastri, Brucks, Kaufman, Cottrell, & Addanki, 1983). The simulator allows the user to design different types of units. These can have any number of input sites, each with an associated site function. Units also have an associated potential and output function. As well as unit types, ISCON allows the user to design different types of weighted link. A network is constructed by generating units of various types and connecting them up. Processing is initiated by activating designated input units. The simulator is implemented on a Prime 550. A network of about 50 units and 300 links takes approximately 30 CPU seconds per iteration. As the number of units increases the simulator takes exponentially longer, making it very unwieldy for networks of more than 100 units. One solution to the speed problem is to compile the networks so that they can be executed faster. A more radical solution, and one which we are currently working on, is to develop a programming language which has as its basic unit a network. This language would involve a batch system rather than an interactive one. There would, therefore, be a trade-off between the ease of use of an interactive system and the speed and power of a batch approach. Although ISCON is an excellent medium for the construction of networks, it is inadequate for any form of sophisticated execution of networks. The proposed Network Programming Language (NPL) would permit the definition and construction of networks in much the same way as ISCON. However, with NPL it will also be possible to selectively activate sections of a particular network, to create new networks by combining separate sub-networks, to calculate summary indices of any network, and to use these indices in guiding the flow of control in the

program. NPL will have a number of modern flow of control facilities (for example, FOR and WHILE loops). Unfortunately, this language is still at the design stage and is not available for use.

IV   THE MODEL

The model consists of five main components which interact in the manner illustrated in Figure 1. The lines ending in filled circles indicate inhibitory connections, the ordinary lines, excitatory ones. Each component consists of sets of neuron-like units which can either excite or inhibit neighbouring nodes, and nodes in connected components. A successful parsing of a sentence is deemed to have taken place if, during the processing of the discourse, the focus is accurately followed, and if at its end there is a stable coalition of only those units central to the discourse. A set of units is deemed a stable coalition if their level of activity is above threshold and non-decreasing.
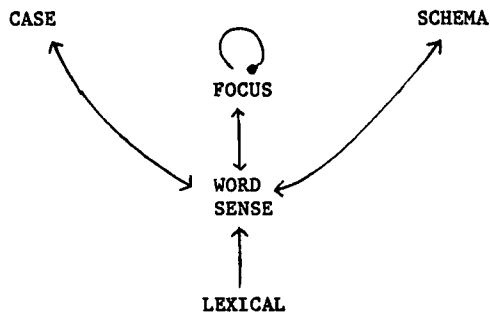

Figure 1. The main components of the model.

A.   Lexical Level

There is one unit at the lexical level for every word in the model's lexicon. Most of the units are connected to the word sense level by unidirectional links, and after activation they decay rapidly. Units which do not have a word sense representation, such as function words and pronouns, are connected by unidirectional link to the case and schema levels. A lexical unit is connected to all the possible senses of the word. These connections are weighted according to the frequency of occurence of the senses. To simulate hearing or reading a sentence the lexical units are activated one after another from left to right, in the order they occur in the sentence.

B.   Word Sense Level

The units at this level represent the "meaning" of the morphemes in the sentence. Ambiguous words are connected to all their possible meaning units, which are connected to each other by inhibitory links. As Cottrell and Small (1983) have shown, this arrangement provides an accuraate model of the processes involved in word sense disambiguation. Grammatical morphemes, function words, and pronouns do not have explicit representations at this level, rather they connect directly to the case and schema levels.

C.   Focus Level

The units at this level represent possible focuses of the discourse in the sense that Sidner (1983) intends. The focus with the strongest activation inhibits competeing focuses. At any one time there is a single dominant focus, though it may shift as the discourse progresses. A shift in focus occurs when evidence for the new focus pushes its level of activation above that of the old one. In keeping with Sidner's (1983) position there are two types of focus used in this model, an actor focus and a discourse focus. The actor focus represents the animate object in the agent case in the most recent sentence. The discourse focus is, as its name suggests, the central theme of the discourse. The actor focus and discourse focus can be one and the same.

D.   Case Level

This model employs what Cottrell and Small (1982) call an "exploded case" representation. Instead of general cases such as Agent, Object, Patient, and so on, more specific case categories are used. For instance, the sentence John kicked the ball would activate the specific cases of Kick-agent and Kick-object. The units at this level only fire when there is evidence from the predicate and at least one filler. Their output then goes to the appropriate units at the focus level. In the example above, the predicate for Kick-agent is kick, and its filler is John. The unit Kick-agent then activates the actor focus unit for John.

E.   Schema Level

This model employs a partial implementation of Small's (1982) proposal for an exploded system of schemas. The schema level consists of a hierarchy of ever more abstract schemas. At the bottom of the hierarchy there are schemas which are so specifc that the number of possible options for filling their slots is highly

constrained, and the activation of each schema serves, in turn, to activate all its slot fillers. Levels further up in the hierarchy contain more general schema details, and the connections between slots and their potential fillers are less strong.

V THE MODEL'S PERFORMANCE

At its current stage of development the model can handle discourse involving pronoun anaphora in which the discourse focus is made to shift. It can resolve the type of reference involved in the following two discourse examples (based on examples by Sidner, 1983; p. 276):

D1-1: I've arranged a meeting with Mick and Peter.

    2: It should be in the afternoon.

    3: We can meet in my office.

    4: Invite Pat to come too.

D2-1: I've arranged a meeting with Mick, Peter, and Pat.

    2: It should be in the afternoon.

    3: We can meet in my office.

    4: It's kind of small,

    5: but we'll only need it for an hour.

In discourse D1, the focus throughout is the meeting mentioned in D1-1. The it in D1-2 can be seen to co-specify the focus. In order to determine this a human listner must use their knowledge that meetings have times, among other things. Although no mention is made of the meeting in D1-3 to D1-4 human listners can interpret the sentences as being consistent with a meeting focus. In the discourse D2 the initial focus is the meeting, but at D2-4 the focus has clearly shifted to my office, and remains there until the end of the discourse.

The network which handles this discourse does not parse it in its entirety. The aim is not for completeness, but to illustrate the operation of the schema level of the model, and to show how it aids in determining the focus of the discourse. Initially, in analyzing D1 the word meeting activates the schema WORK_PLACE_MEETING. This schema gets activated, rather than any other meeting schema, because the overall context of the discourse is that of an office memo. Below, is a representation of the schema. On the left are its component slots, and on the right are all the possible fillers for these slots.

WORK_PLACE_MEETING schema

WPM_location: library
                           tom_office
                           my_office

WPM_time: morning
                           afternoon

WPM_participants: tom
                           vincent
                           patricia
                           mick
                           peter
                           me

When this schema is activated the slots become active, and generate a low level of subthreshold activity in their potential fillers. When one or more fillers become active, as they do when the words Mick and Peter are encountered at the end of D1-1, the slot forms a feedback loop with the fillers which lasts until the activity of the sense representation of meeting declines below a threshold. A slot can only be active if the word activating the schema is active, which in this case is meeting. When a number of fillers can fill a slot, as is the case with the WPM_participant slot, a form of regulated sub-network is used. On the other hand, when there can only be one filler for a slot, as with the WPM_location slot, a winner-take-all network is used (both these types of sub-network are described in Feldman and Ballard, 1982).

Associated with each unit at the sense level is a focus unit. A focus unit is connected to its corresponding sense unit by a bidirectional excitatory link, and to other focus units by inhibitory links. As mentioned above, there are two separate networks of focus units, corresponding to actor focuses and discourse focuses, respectively. Actors are animate objects which can serve as agents for verbs. An actor focus unit can only become active if its associated sense level unit is a filler for an agent case slot. The discourse focus and actor focus can be, but need not be, one and the same. The distinction between the two types of focus is in line with a similar distinction made by Sidner (1983). The structure of the focus level network ensures that there can only be one discourse focus and one actor focus at a given time. In discourses D1 and D2 the actor focus throughout is the speaker.

At the end of the sentence D1-1 the WORK_PLACE_MEETING schema is in a stable coalition with the sense units representing Mick and Peter. The focus units active at this stage are those representing the speaker of the discourse (the actor focus), and the meeting (the discourse focus). When the sentence D1-2 is

encountered the system must determine the co-specification of it. The lexical unit it is connected to all focus units of inanimate objects. It serves to boost the potential of all the focus units active at the time. At this stage, if there are a number of competitors for co-specification, a number of focus units will be activated. However, by the end of the sentence, if the discourse is coherent, one or other of the focuses should have received sufficient activation to suppress the activation of its competitors. In the case of D1 there is no competitor for the focus, so the it serves to further activate the meeting focus, and does so right from the beginning of the sentence.

The sentence D1-3 serves to fill the WPM_location slot. The stable coalition is then enlarged to include the sense unit my office. The activation of my office activates a schema, which might look like this:

MY_OFFICE schema

MO_location:	Prefab 1
MO_size:	small
MO_windows:	two

It is not strictly correct to call the above structure a schema. Being so specific, there are only single fillers for any of its slots. It is really a representation of the properties of a specific office, rather than predictions concerning offices in general. However, in the context of this type of model, with the emphasis on highly specific rather than general structures, the differences between the two schemas presented above is not a clearcut one.

When my office is activated, its focus unit also receives some activation. This is not enough to switch the focus away from meeting. However, it is enough to make it a focus candidate, which would permit a switch in focus in the very next sentence. If a switch does not take place, the candidate's level of activity rapidly decays. This is what happens in D1-4, where the sentence specifies another participant, and the focus stays with meeting. The final result of the analysis of discourse D1 is a stable coalition of the elements of the WORK_PLACE_MEETING frame, and the various participants, times, and locations mentioned in the discourse. The final actor focus is the speaker, and the final discourse focus is the meeting.

The analysis of discourse D2 proceeds identically up to D2-4, where the focus shifts from meeting to my office. At the beginning of D2-4 there are two candidates for the discourse focus, meeting and my office. The occurence of

the word it then causes both these focuses to become equally active. This situation reflects our intuitions that at this stage in the sentence the co-specifier of it is ambiguous. However, the occurence of the word small causes a stable coalition to form with the MY_OFFICE schema, and gives the my office focus the extra activation it needs to overcome the competing meeting focus. Thus, by the end of the sentence, the focus has shifted from meeting to my office. By the time the it in the final sentence is encountered, there is no competing focus, and the anaphor is resolved immediately.

There are a number of fairly obvious drawbacks with the above model. The most important of these being the specificity of the the schema representations. There is no obvious way of implementing a system of variable binding, where a general schema can be used, and various fillers can be bound to, and unbound from, the slots. It is not possible to have such symbol passing in a connectionist network. Instead, all possible slot fillers must be already bound to their slots, and selectively activated when needed. To make this selective activation less unwieldy, a logical step is to use a large number of very specific schemas, rather than a few general ones.

Another drawback of the model proposed here is that there is no obvious way of showing how new schemas might be developed, or how existing ones might be modified. One of the basic rules in building connectionist models is that the connections themselves cannot be modified, although their associated weights can be. This means that any new knowledge must be incorporated in an old structure by changing the weights on the connections between the old structure and the new knowledge. This also implies that the new and old elements must already be connected up. In spite of the apparent oversupply of neuronal elements in the human cortex, to have everything connected to virtually everything else seems to be profligate.

Another problem with connectionist models is their potential "brittleness". When trying to program a network to behave in a particular way, it is difficult to resist the urge to patch in arbitrary fixes here and there. There are, as yet, no equivalents of structured programming techniques for networks. However, there are some hopeful signs that researchers are identifying basic network types whose behavior is robust over a range of conditions. In particular, there are the winner-take-all and regulated networks. The latter type, permits the specification of upper and lower bounds on the activity of a sub-network, which allows the designer to avoid the twin perils of total saturation of the network on the one hand, and total silence on the other. A reliable taxonomy of sub-networks would greatly aid the designer in building robust networks.

VI   CONCLUSION

This paper briefly described the connectionist approach to cognitive modelling, and showed how it might be applied to langauge processing. A connectionist model of language processing was outlined, which employed schemas and focusing techniques to analyze fragments of discourse. The paper described how the model was successfully able to resolve simple it anaphora.

A tape of the simulator used in this paper, along with a specification of the network used to analyze the sample discourses, is available from the author at the above address, upon receipt of a blank tape.

VII   REFERENCES

Cottrell, G.W., & Small, S.L. (1983). A connectionist scheme for modelling word sense disambiguation. Cognition and Brain Theory, 6, 89-120.

Feldman, J.A., & Ballard, D.H. (1982). Connectionist models and their properties. Cognitive Science, 6, 205-254.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

Sidner, C.L. (1983). Focussing in the comprehension of definite anaphora. In M. Brady & R.C. Berwick (Eds.), Computational models of discourse, Cambridge, Massachusetts: MIT Press.

Small, S.L. (1982). Exploded connections: Unchunking schematic knowledge. In Proceedings of the Fourth Annual Conference of the Cognitive Science Society, Ann Arbor, Michigan.

Small, S.L., Cottrell, G.W., & Shastri, L. (1982). Toward connectionist parsing. In Proceedings of the National Conference on Artificial Intelligence, Pittsburgh, Pennsylvania.

Small, S.L., Shastri, L., Brucks, M.L., Kaufman, S.G., Cottrell, G.W., & Addanki, S. (1983). ISCON: a network construction aid and simulator for connectionist models. TR109. Department of Computer Science, University of Rochester.