A sensorimotor interpretation of Logical Form, and its application in a model of Māori sentences

Alistair Knott and Martin Takac

1. Introduction

In this paper we will present a model of the syntax of Māori transitive clauses. In one sense, the model will be familiar to linguists working within the Government-Binding/Minimalist tradition, but in another sense, it will be novel, because it is implemented as a neural network, within a model of nonlinguistic sensory and motor processing.

The guiding hypothesis behind the model is that the logical form (LF) of a sentence reporting some directly observable event in the world can be interpreted as a description of the sensory and/or motor processes involved in *experiencing* this event. This proposal is described in more detail elsewhere (see e.g. Knott 2012; Takac et al. 2012). We hypothesise that language is intimately connected with the sensorimotor mechanisms through which we experience the world. This idea has received a lot of attention in cognitive science, within 'embodied' models of cognition (see e.g. Feldman & Narayanan 2004; Barsalou 2008). But the idea has some interesting implications about language universals that cognitive scientists do not typically pursue. If language is connected to sensorimotor mechanisms, then we expect structural similarities between all languages, because speakers of all languages have the same sensorimotor mechanisms. If language is *strongly* connected to sensorimotor mechanisms, as many embodied linguists believe, then we should expect a substantial set of structural similarities between languages. Such similarities are clearly not visible 'on the surface', so the only way to maintain a strongly embodied model of language is to adopt some linguistic theory that posits cross-linguistic universals at some 'underlying' level of structural representation. This argument provides an interesting way of thinking about Chomskyan models of syntax. Chomskyan models take linguistic universals seriously: identifying underlying structures that obtain in many languages is at the heart of the Chomsykan research programme. From this perspective, a Chomskyan account of syntax might provide an ideal vehicle for the expression of 'strongly' embodied models of language.

This suggestion upsets Chomskyan linguists and cognitive scientists in equal measure. Cognitive scientists tend not to like Chomskyan models – they are not implemented, they provide no account of sentence processing, and they cannot represent the collocational surface structures in text that modern statistical linguistics is so good at characterising. Chomskyan linguists often see the project of looking for neural correlates of syntactic structure as peripheral to the main work to be done. Liz, it must be said, has been quite supportive of the line of work we are pursuing – though she did say she was glad she's not the one doing it!

We'll begin in Section 2 by sketching a simple LF template for a transitive sentence. In Section 3 we will outline a sensorimotor interpretation for this LF structure. The basic idea is that the LF structure describes the process of 'rehearsing' a sensorimotor process – in this case, the process of perceiving an event involving a transitive action. In Section 4 we will describe a neural network mechanism that implements this sensorimotor rehearsal process. This mechanism doubles up as a sentence generator: during rehearsal, sensorimotor representations that become active can trigger output phonological representations, through a network that is trained by exposure to a particular language. The training process involves learning the meanings of individual word stems and inflections, but also involves a process akin to parameter-setting: the network has several opportunities to generate phonological signals reflecting the semantic constituents of a transitive sentence, and learns to take the opportunities that result in surface structures in the exposure language. Our hope is that when Chomskyan linguists look at this network, and screw up their eyes a bit, they can see its sensorimotor representations as encoding LF structures, and the mechanism that maps sensorimotor representations onto output phonology as a device for learning the parameters that map LF to PF in a particular language. In Section 5 we will show how the network can learn some simple Māori sentences.

2. LF structure of a transitive sentence

The LF structure we assume for a transitive sentence is shown in Figure 1.

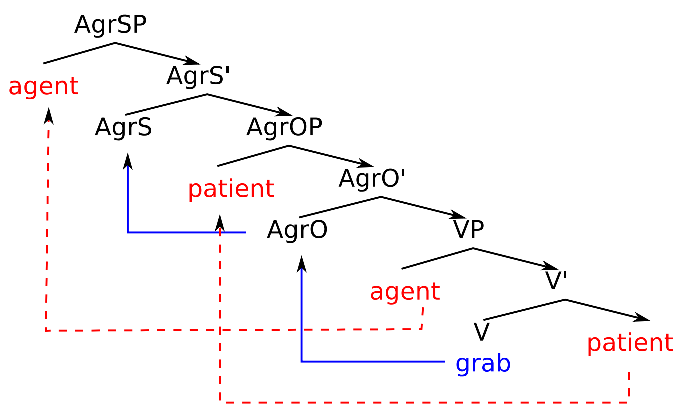

Figure 1: Schematic LF structure of a transitive clause

This corresponds roughly to the structure sketched by Chomsky (1995), summarising the GB model as it had advanced to by that time, in preparation for his initial statement of the Minimalist model. It incorporates Pollock's (1989) suggestion that the agreement features of the verb occupy a separate functional projection above VP, and Koopman & Sportiche's (1991) suggestion that the subject of a sentence is base-generated at the specifier of VP. Chomsky's (1995) analysis added the suggestion that there are two agreement projections in a transitive clause, one for the subject and one for the object. In this model, both the subject and the object of a transitive clause are base-generated in VP: the subject in [Spec,VP] and the object in [Comp,V]. They each raise to the specifier of a higher functional projection, to be assigned Case: the subject to [Spec,AgrSP] and the object to [Spec,AgrOP] (see the red arcs in the figure). The verb raises by head-movement successively to the heads of AgrO and AgrS (see the blue arcs in the figure). The justification for the movement of the verb is easier to give in Minimalist terms. The verb is fully inflected when generated in the V head, and has to raise to these two heads to 'check' its agreement features. For now, we will omit the tense projection that featured in Chomsky's model, but we will introduce it in Section 5.1.

This model is attractive because it simplifies both Case-assignment mechanisms and theta-role-assignment mechanisms. Chomsky's positing of two agreement projections simplifies Case assignment, because Case is now uniformly assigned by a functional head to its specifier. Koopman & Sportiche's positing of a VP-internal subject simplifies theta-role assignment, because 'agent' and 'patient' roles (or 'proto-agent' and 'proto-patient', to use Dowty's 1991 terminology) can now be assigned by the verb locally, within its maximal projection, to its specifier and complement positions respectively.

The structure sketched in Figure 1 provides the basis for a simple account of Māori transitive sentences. To account for the VSO structure that is typical of these sentences, we can posit that in Māori, V raises to its high position before spell-out, while S and O raise to their Case-assigning positions after spell-out, so that at PF, V is pronounced at its 'high' position (in our sketch, at the AgrS head), while S and O are pronounced at their base positions in VP. This account of VSO languages was one of the motivations for Koopman & Sportiche's model of VP-internal subjects, and several models of Māori along these lines have been developed, among which Liz's models feature prominently (see e.g. Pearce & Waite 1997; Pearce 2000). More recent models of Māori sometimes extend or revise this scheme (e.g. Pearce 1998; 2002), but for our account we will adopt this Chomskyan model, preserved in aspic from 1995.

3. A sensorimotor interpretation of the LF structure of a transitive clause

All linguists think of syntactic structures as having cognitive significance: they portray something about how sentences are represented in the brain. How does the LF structure sketched in Figure 1 do this? One suggestion is that it somehow describes a cognitive *representation*: something stored in a pattern of activity somewhere in the brain, or in a pattern of synaptic connections. Our suggestion is that it describes a *process* that takes place in the brain, rather than a static representation. Specifically, it describes a process whereby a particular sensorimotor experience is *rehearsed*, or relived. We assume a particular model of sensorimotor processing, which emphasises the *sequential structure* of the sensory and motor operations through which we interact with the world. The basic principles of this model were introduced by Ballard et al. (1997). As set out persuasively in that paper, sensorimotor operations often have to be executed in a particular sequence: for instance, an agent cannot readily classify an object presented visually until she has attended to it, overtly or covertly, and cannot reach for a target object until it has been both attended to and classified.

3.1. A sensorimotor model of reaching-to-grasp

Drawing on a large body of experiments in neuroscience, we have developed a model of the sequence of sensorimotor processes that an observer must execute in order to experience an event involving a simple transitive action – a reach-to-grasp action (see Knott 2012 for details). Following Ballard et al., we hypothesise that the atomic elements of this sequence are all operations of the same basic type: a sensory or motor operation is executed (which we term a **deictic operation**), which updates the observer's current physical and cognitive state (which we term the observer's **context**), generating a sensory representation as a side-effect (which we term the **reafferent signal**). The new context permits the execution of other deictic operations; thus sensorimotor processing is naturally structured into *sequences* of deictic operations. We call these sequences **deictic routines**, again following Ballard et al.

The deictic routine involved in experiencing a reach-to-grasp action is illustrated in Figure 2.

Context	Deictic operation	Reafferent signal
C1	Attend to agent	Attending-agent
C2	Attend to patient	Attending-patient
C3	Execute motor action	Reattending-agent
C4		Reattending-patient

Figure 2: The deictic routine involved in experiencing a reach-to-grasp action

It comprises three deictic operations. The first operation is an action of attention to the agent. This could be implemented in an operation like a saccade, that points the observer's fovea towards a particular external agent in the world. But it could also be a more internal action of attention that focuses the observer's attention on herself: this is what happens when the observer 'decides to act', thereby selecting herself as the agent of whatever action takes place next. In each case, the attentional action allows activation of a representation of the agent as a reafferent consequence.

The second operation is an action of attention to the target of the reach action. If the observer is the agent, this involves directing attention to an object in her peripersonal space. If the agent is some external actor, it involves following the gaze of this actor to identify the intended target of her reach action. Again, in either case, the attentional action allows activation of a representation of the target object as a reafferent consequence.

The third operation is one whereby the observer monitors a continuous motor action in real time, until it is completed. Interestingly, during this process, the observer activates the category of the action in question, but also activates a second representation *of the agent*, as a reafferent consequence of action monitoring. This time the agent is represented as a dynamic, articulated entity, rather than just as a static object of attention. At the completion of the monitored action, the observer also activates a second representation *of the target object*, again within the motor modality: roughly speaking, the location of the object is represented by the location of the agent's arm, and the shape of the object is represented by the shape of the agent's hand.

3.2. LF structure as a representation of a rehearsed deictic routine

There are many similarities between the structure of the deictic routine for transitive actions sketched in Section 3.1 and the LF structure of a transitive sentence outlined in Section 2. In each case, the structure is composed of instances of a recursively defined 'basic building block'. For LF structure, the building block is the X-bar schema, and the recursive principle is the one which allows an XP to occupy the complement of another XP. For the deictic routine, the building blocks are deictic operations. These building blocks also align well with each other: in each case, we have an element associated with the (proto-)agent, followed by an element associated with the (proto-)patient, followed by an element associated with the action. Finally, in each case, there are two representations of the agent, and two representations of the patient.

These similarities suggest an interesting cognitive interpretation of LF structure. According to this interpretation, the LF structure of a sentence reporting a transitive event represents the deictic routine through which this event was experienced. Each XP in the right-branching LF structure identifies one of the deictic operations in the routine. The head of each XP denotes a deictic operation, while its specifier denotes its reafferent consequence. The right-branching structure of XPs identifies the sequential order in which the deictic operations occur. This interpretation is illustrated in Figure 3.

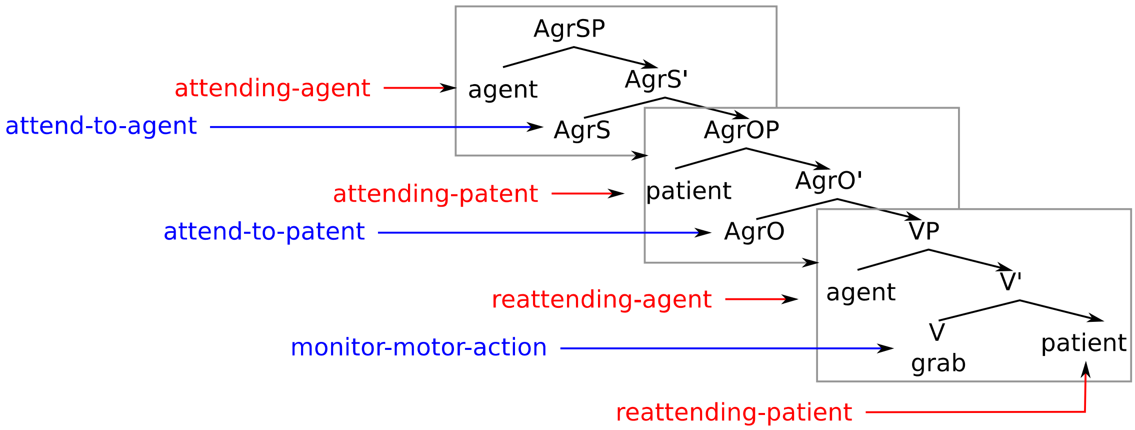

Figure 3: A sensorimotor interpretation of the LF of a transitive clause

This way of thinking about LF requires some mental adjustments. We now have to think about an LF structure not as the product of some abstract process of derivation, but as a reflection of an *actual* cognitive process, that can be directly studied. (What is more, this process is not a linguistic process *per se*, but a sensorimotor one.) Moreover, the operations that 'move' constituents from one LF position to another now have a completely different significance. These movements now have a *temporal* interpretation. For instance, when we see the subject DP raising from a VP-internal position to a higher position, we have to interpret this as implying that there are two *times* when the subject 'appears'. Similarly, the raising of the object DP tells us there are two *times* when the object 'appears'.

However, we suggest that thinking of LF structures as representing processes in this way can be extremely useful. For instance, consider the topic of DP-raising. The requirement that DPs raise 'to get Case' is ultimately motivated because it contributes to a descriptively adequate and economical model of many languages. But it would be nice to justify it in more concrete terms as well. In the sensorimotor interpretation of LF that we propose, the raising of DPs is a manifestation of a constraint *on sensorimotor processing*: an observer has to *attend* to the agent and patient of a transitive action (in that order) before she can monitor this action (both for actions she executes herself, and for actions she perceives being executed by other agents). Thinking of an LF structure as representing a sensorimotor sequence is also helpful from the perspective of linking models of LF derivation to models of sentence processing. In the interpretation we suggest, an LF structure does not just represent a speaker's declarative knowledge of language – it *directly represents a cognitive process* – and moreover, one which is plausibly involved in the actual generation of sentences. We will flesh this idea out in Section 4.

The idea that LF structure encodes a sequence is not completely foreign to linguists. Kayne's (1994) model of LF, which was another of the influences in Chomsky's (1995) model, stipulates that the specifier of an XP appears before its complement at PF. Kayne does also explicitly state that LF has hierarchical, and not 'linear' (i.e. temporal) structure. However, it has at least an *implicit* temporal structure, in the structure it imposes on PF. And in fact, he tangentially suggests that the right-branching form of LF structures in his model may have a temporal origin.

Before we introduce our network model, we need to consider what head movement means in our reinterpreted conception of LF. Head movement allows a fully inflected verb to 'raise' from the head of V, through the head of AgrO, to the head of AgrS. If the head of each XP denotes a deictic operation, and the right-branching structure of these XPs denotes the sequence in which they occur, then the mechanism of head movement allows for deictic operations to be reported '*out-of-sequence*': the motor action denoted by the V head appears 'too soon' when the inflected verb occupies its 'high' positions, and the actions of attention to agent and patient denoted by the Agr heads appear 'too late' when the inflected verb occupies its 'base' position. In our interpretation of this phenomenon, we introduce another component to the sensorimotor model. Sentences are not generated as a direct side-effect of sensorimotor experience: rather, they are produced from a representation of an experienced event held in *working memory*. Our model of sensorimotor processing includes a model of working memory for experienced events. In this model, an agent stores an experienced event as a *prepared deictic routine* that can be *replayed*. In particular, it can be replayed in a special mode, in which activated *sensorimotor* representations can trigger output *phonological* representations. A great deal is known about how deictic routines are stored in the brain. Crucially, the working memory representation of a deictic routine holds representations of all its component operations active *in parallel*, even though they are executed sequentially (see for instance Averbeck et al. 2002). We propose that the phenomenon of head-raising arises because heads are phonological expressions of deictic operations *as represented in working memory*, rather than in the sensorimotor media where they occur transiently during actual sensorimotor experience. We propose that the LF of a sentence describes a deictic routine *replayed from working memory*, rather than one occurring in real time. During this process, there is an interesting mixture of 'sustained' and 'transient' representations. We propose that heads are read from the 'sustained' ones, and specifiers are read from the 'transient' ones.

With these preliminaries, we can now introduce our model of sentence generation.

4. A neural network model of sentence generation

Our sensorimotor conception of LF lends itself to a model of sentence processing – specifically, a model of sentence generation. As just described, we think of an LF structure as a representation of the mixture of sustained and transient sensorimotor representations that are activated when a deictic routine, encoding a recently experienced event, is replayed from working memory. We envisage that this replay operation can happen in a special mode, in which active sensorimotor representations trigger phonological representations: in this mode, a sequence of phonological representations will be produced. Thus we see the process of sentence generation as a process that maps a rehearsed deictic routine onto a phonological sequence.

In this section, we will introduce a computational model of this generation process. It is implemented as a neural network. For details about the architecture of the network and its training, see Takac et al. (2012).

The basic structure of the network is shown in Figure 4. It takes a sequence of inputs, at three successive time steps, and at each step has the opportunity to generate a phonological output. Its inputs come both from the 'sustained' representations of the complete deictic routine held in working memory, which are the same at each time step, and from the 'transient' representations of individual operations in the routine that change at each step. At each step, there is a mechanism that selects first the transient representation and then the sustained one.

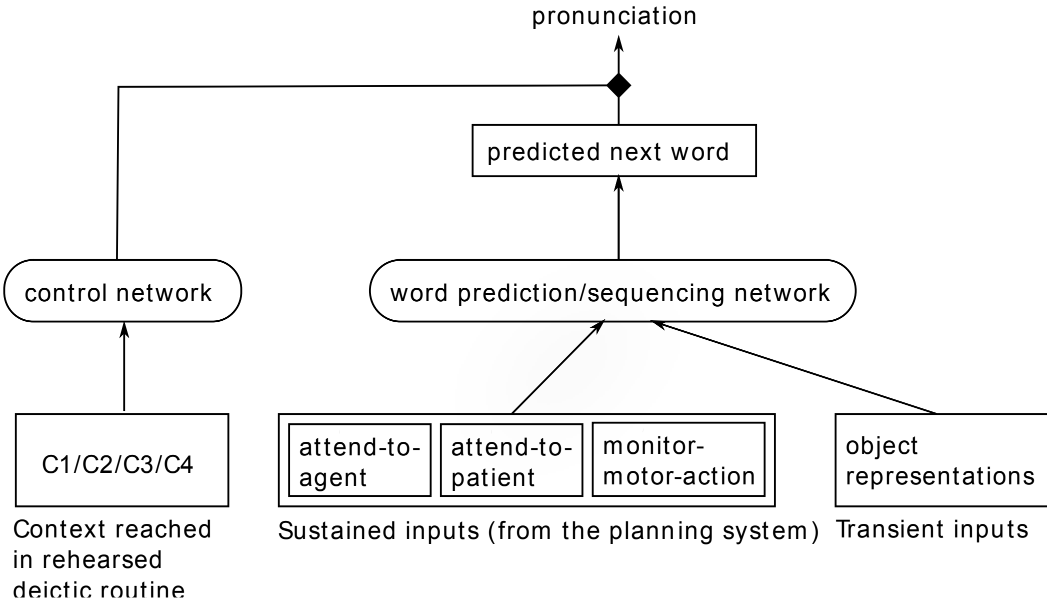

Figure 4: Architecture of the sentence generation network

There are two sub-networks. The **word production and sequencing network** (or WPSN) takes the currently selected sensorimotor input representation, and maps it onto a predicted output phonological representation (a stem and an inflection). The **control network** decides whether or not this phonological representation should be explicitly *pronounced*. It does this using information about what 'stage' of rehearsal has been reached: this is an encoding of the 'context' that is updated after each sensorimotor operation.

Both networks are trained on sentences from a given language that denote concrete events. Each sentence is paired with the deictic routine through which the associated event is experienced, according to the model outlined in Section 3.1. At the start of training, we assume the learner's ability to rehearse a deictic routine is poor, and items from the routine are paired indiscriminately with words in the associated sentence. During this time, the WPSN slowly learns a small set of word meanings – that is, associations between sensorimotor signals and output word stems – through a process called 'cross-situational learning' (Siskind 1996). At a certain point, the learner becomes able to rehearse deictic routines accurately. At this point, the control network starts to be trained in addition. Training now involves rehearsing each deictic routine in its proper sequence, to produce a sequence of output words: these words are compared to the words in the associated training sentence, beginning with the first word.

The WPSN is trained to produce the 'current word' in the training sentence. Meanwhile, the control network is trained in the meta-level task of when to overtly *pronounce* the words produced by the WPSN. As already mentioned, a rehearsed deictic routine provides several opportunities to produce the key constituents in a transitive sentence. There is an 'early' opportunity to produce a word denoting the agent in the first rehearsed deictic operation (attention to the agent), and then a 'late' opportunity to produce such a word in the third rehearsed operation (monitoring of the motor routine, where the agent's characteristic pattern of movement is represented). Similarly, there is an 'early' opportunity to produce a word denoting the patient in the second rehearsed operation (attention to the patient), and a 'late' opportunity to pronounce such a word in the third rehearsed operation (where the patient is represented as a motor state of the agent's hand/arm). Finally, there are several opportunities to produce phonological outputs denoting the deictic operations themselves, because these outputs are generated from tonic representations of these operations in the planning system, which are active throughout the rehearsed routine. From all these opportunities to pronounce words, the control network learns to take opportunities that result in surface sentence forms resembling those of the training language.

Consider a schematic Māori training sentence, comprising a verb, a subject and an object, in that order, and the deictic routine paired with this sentence in training. The network receives each operation in the rehearsed deictic routine in turn. In the first operation, it receives first the agent (an opportunity to pronounce the subject), and then the set of planned deictic operations (an opportunity to pronounce the verb and associated inflections). Assume the WPSN correctly generates the word denoting the agent. This word is compared to the first word in the training sentence – because they are not the same, the control network will learn (incrementally) that it should *not pronounce* the agent in this early 'context'. Assume the WPSN also correctly generates a word denoting the action. This word is compared to the first word in the training sentence – this time it does match, and the control network learns (incrementally) that it *should pronounce* the action in this early 'context'. In syntactic terminology, the training sentence provides a small piece of evidence in favour of pronouncing verbs/inflections 'high' in Māori, and against pronouncing subjects 'high'.

The above example assumes that the WPSN has already learned the word forms denoting the relevant sensorimotor symbols. However, this is not always the case: the WPSN still has to learn many words. But as the control network learns the right 'opportunities' to pronounce words in the training language, it also generates improved training data for the WPSN. When the control network learns not to pronounce the subject 'high' in Māori, it refrains from training the WPSN to map the agent onto the first word in a Māori sentence. As the WPSN's learning of word meanings improves, it in turn generates cleaner training data for the control network, so the two networks 'bootstrap' each other. This simulates the effect whereby knowledge of syntax aids word learning (see e.g. Aslin et al. 1996).

There is one other important feature of the WPSN to introduce. This network takes sensorimotor signals and learns to generate word forms, as already noted. But it also maintains a record of the sequence of words *produced so far* in the sentence being generated: its decision about how to map sensorimotor signals onto word forms is conditioned on this sequence of recent words. This provides the network with a mechanism for producing sequences of words that conform to surface regularities in the exposure language. It has the ability to learn 'idiomatic' or collocational structures in language – an ability that is hard to model within a traditional GB or Minimalist paradigm (as notoriously discussed by Jackendoff (2002) in his criticism of the paradigm). We see this ability to learn idioms as one of the key benefits of our proposed reinterpretation of LF structures. The model outlined in this section is a model of how LF structures participate in actual sentence processing – that is, in routines that generate surface sentence forms. Within this model, we can introduce machinery that learns regularities in surface sentence structure, *in addition to* the machinery that learns traditional GB/Minimalist-style parameter settings. We hope that a Chomskyan linguist will be able to look at our model (with eyes screwed up) and see LF structures, head movement, DP-raising, and parameter-setting. But an empiricist linguist should also be able to look at our network and see standard neural network mechanisms for learning surface collocations and idioms.

5. Training the network on a corpus of Māori sentences

5.1. Adding tense and causative actions to the deictic routine

One of the distinctive features of Māori (along with other Polynesian languages) is its use of tense markers. In Chomsky's (1995) model, tense information is contributed by a separate functional projection, 'high' in the LF structure: we will assume the tense projection TP is the highest projection, above AgrSP and AgrOP. Importantly, we have to extend our sensorimotor model of transitive events to incorporate a deictic operation that occurs before 'attention to the agent', which plausibly contributes tense information. Our suggestion is that this operation is one which determines whether the observer attends to the perceptual here-and-now as a source of incoming events, or to his own episodic memory. The operation of attending to the here-and-now corresponds to a 'present-tense' head; the operation of 'engaging episodic memory' corresponds to a 'past-tense' head. We assume a particular neural network model of episodic memory, presented elsewhere (Takac & Knott 2016a; 2016b), in which events are stored in and retrieved from long-term memory in the form of deictic routines, with the same structure as those generated during experience. In this model, events are retrieved from episodic memory *into working memory*, from where they can be rehearsed like events that have just been perceived.

Another distinctive feature of Māori is its productive use of the causative prefix *whaka* on verbs. Again, we can extend the LF structure of the clause to model this. Our model of causatives, like many others, is based on Larson's (1988) concept of VP shells: we assume an 'outer' VP headed by 'cause' that introduces an 'inner' VP denoting the caused event. Again, we must extend the sensorimotor model, to provide a plausible sensorimotor correlate for the outer VP, and its relationship to the inner VP. We have developed a neural network model of causative actions, again presented elsewhere (Lee-Hand & Knott 2015) in which there are correlates both of the causative action and of the caused event.

With these preliminaries, we can introduce an experiment in which our sentence generation network was trained on a corpus of Māori sentences.

5.2. A training corpus of Māori sentences

Our training corpus consisted of 160,000 sentences (16 training epochs each with 10,000 sentences) generated at random from a simple grammar, out of which approximately 60% were transitive sentences, 27% causative sentences and 13% intransitive sentences. Intransitive sentences lacked the AgrO projection, and had no object. Each sentence had a tense/aspect marker, which was either *i* (for past tense) or *kei te* (for present continuous) or *e (…) ana* (for continuous aspect, with tense unspecified). The parentheses in *e (…) ana* indicate the position of the verb. We did not include a separate projection for aspect (this is a topic for ongoing work). Instead, our network was required to learn to produce *e (…) ana* from the same deictic routines as *i*. We used a range of DPs that were semantically suited to the argument roles of verbs. DPs could include pronouns (first/second/third-person and singular/dual/plural); dual and plural first-person pronouns could be exclusive or inclusive. We also included reflexive pronouns. Third-person DPs using common nouns could use definite or indefinite determiners. Common nouns could be singular or plural; we included some irregular plural nouns (*tīpuna* 'ancestors, grandparents', *wāhine* 'women', *tamariki* 'children'). We used 31 open-class verbs in our example sentences, and 42 open-class nouns. Intransitive verbs could participate in causative constructions; in that case, for technical reasons, the causative prefix *whaka* appeared as a separate word. Object DPs were introduced with the particle *i*: again, this particle had to be learned as an idiom. Finally, we included some additional continuous idioms in the training sentences (*kai moana* 'seafood, shellfish', *pene rākau* 'pencil', *tipuna whaea* 'great grandmother', *tipuna matua* 'great grandfather', *taonga tākaro* 'traditional games'). In addition, since we had not provided a dedicated sensorimotor operation for *ana* in the *e (…) ana* construction, this construction also functions as an 'idiom' for our network: in this case, a discontinuous one. And since our deictic routines do not model the internal structure of DPs, determiner-noun constructs also function as (continuous) idioms for our network – as they apparently do for infants at a certain developmental stage (see e.g. Pine and Lieven 1997).

Some examples of the training sentences in our corpus are given below.

1. E whaka hoki ana kōrua i te parāoa.
CONT CAUS return CONT 2DU OBJ DET.SG bread
'You [dual] are returning the bread.'

2. I mātakitaki tāua i ngā taonga tākaro.
PAST watch 1DU.INCL OBJ DET.PL games
'We [dual, inclusive] watched the games.'

3. Kei te horoi tātou i a kōrua.
PRES wash 1PL.INCL OBJ PERS 2DU
'We [plural, inclusive] wash you [dual].'

4. Kei te whaka ngaro ahau i a māua.
PRES CAUS be.hidden 1SG OBJ PERS 1DU.EXCL
'I hide us [dual, exclusive].'

5. | E | whaka | makere | ana | he | tamaiti | i | ngā | kau. |
|---|---|---|---|---|---|---|---|---|
| CONT | CAUS | fall | CONT | INDEF | child | OBJ | DET.PL | cow |

‘A child drops the cows.’

5.3. Results

In each training run, we trained our network on 10,000 sentences of the kind described above, each paired with its associated deictic routine. To assess the network’s performance, we tested it by presenting it with the deictic routines associated with each of the 10,000 training sentences and the deictic routines associated with an additional 1,000 sentences unseen during training, and asking it to generate a sentence from each. We compared the generated sentence to the sentence paired with the deictic routine during training. The sentence was judged to be ‘correctly’ generated if it matched the paired sentence in every respect, modulo synonyms. The model was able to correctly generate 99.2% of training sentences and 98.5% of unseen ones.

4. Discussion

In this paper we have described a neural network that can learn a fragment of a natural language grammar, when trained on sentences from a given language, paired with semantic representations. The semantic representations we use are distinctive, in that they derive directly from a model of sensorimotor processing, rather than being expressed in an artificial logical language. But they are also distinctive in having a direct correspondence with Chomskyan LF structures. In our model, constraints on the structure of sensorimotor routines, and on the working memory mechanisms that store and replay them, are reflected in the space of possible surface languages: our network makes use of these constraints to learn the grammar of its exposure language. In this model, the innate ‘knowledge of language’ that is captured by LF structure is (at least partly) due to the structure of the sensorimotor system. In this sense our model is an ‘embodied’ model of language. But by the same token, it is also a ‘nativist’ model, of an interesting new kind. At the same time, our sentence generation network can also learn idiomatic surface structures in the exposure language: it thus implements a mixture of nativist and empiricist models.

Note that if the network is exposed to training corpora from other languages, it will learn different parameter settings that choose different positions for verb heads and their arguments. The network has also been trained on SVO languages (English and Slovak) and SOV languages (Japanese), and performs at a similar level. It can learn to express tense in verb inflections as well as in stand-alone particles. It can learn to produce subject and object agreement inflections on verbs, or to omit these. It can learn to produce pronouns as clitics adjoined to verb heads (as in Slovak) or in regular argument positions (as in English). It can learn to realise causative actions with an explicit prefix (as in Māori), or without (as in English). For details of these experiments, see Takac et al. (2012) (plus papers in preparation on Slovak and Japanese). The Māori experiments reported here are particularly useful in demonstrating an ability to learn a rich pronoun paradigm, stand-alone tense markers, and morphology realising the ‘cause’ concept in causative sentences.

Of course, we are just scratching the surface of the complexity of Māori sentence structure. We do not have a model of passive sentences, which often provide the most natural way of rendering events in Māori. We have not begun to model the distinctive topicalising projections in the left periphery, or the internal structure of Māori DPs – or indeed any forms of predicative or stative sentence – all topics that Liz has studied in great depth. Our grammar development methodology is rather slow compared to that of a theoretical linguist: every LF structure has to be justified not only by its role in a wider model of grammar, but also as a plausible deictic routine, motivated by research in neuroscience, and tested in a neural network model. However, we are not deterred by this slower pace of progress: we think it is helpful to use deictic routines to ‘triangulate’ on LF structures in this way. One difficulty with theoretical linguistics in general is that there are often many plausible theoretical analyses of a given phenomenon: the data frequently underdetermine the space of possible theories. If models of sensorimotor processing can provide additional constraints on the process of building syntactic models, that could be a good thing in the long run.

Acknowledgements

The authors would like to acknowledge Marsden grant 13-UOO-048 for supporting this work.

References

Aslin, Richard N. & Woodward, Julide Z. & LaMendola, Nicholas P. & Bever, Thomas G. 1996. Models of word segmentation in fluent speech to infants. In Morgan, James L. & Demuth, Katherine (eds.) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 117–134. Mahwah, NJ: Erlbaum.

Averbeck, Bruno B. & Chafee, Matthew V. & Crowe, David A. & Georgopoulos, Apostolos P. 2002. Parallel processing of serial movements in pre-frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 99(20). 13172–13177.

Ballard, Dana H. & Hayhoe, Mary M. & Pook, Polly K. & Rao, Rajesh P.N. 1997. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences* 20(4). 723–767.

Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology* 59. 617–645.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Feldman, Jerome & Narayanan, Srinivas. 2004. Embodied meaning in a neural theory of language. *Brain and Language* 89(2). 385–392.

Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Knott, Alistair. 2012. *Sensorimotor cognition and natural language syntax*. Cambridge, MA: MIT Press.

Koopman, Hilda & Sportiche, Dominique. 1991. The position of subjects. *Lingua* 85. 211–258.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–391.

Lee-Hand, Jeremy & Knott, Alistair. 2015. A neural network model of causative actions. *Frontiers in Neurorobotics* 9. Article 4.

Pearce, Elizabeth. 1998. Incorporation in Māori syntax. In *Proceedings of AFLA V*, University of Hawai'i.

Pearce, Elizabeth. 2000. Argument positions and anaphora in the Māori clause. In Fisher, Steven Roger & Sperlich, Wolfgang (eds.) *Leo Pasifika: Proceedings of the Fourth International Conference on Oceanic Linguistics*, vol. 2, 313–325. Auckland: The Institute of Polynesian Languages and Literatures.

Pearce, Elizabeth. 2002. VP versus V raising in Māori. In Rackowski, Andrea & Richards, Norvin (eds.) *Proceedings of AFLA VIII: The Eighth Meeting of the Austronesian Formal Linguistics Association*, 225–240. Cambridge, MA: MIT Working Papers in Linguistics.

Pearce, Elizabeth & Waite, Jeffrey. 1997. Kia and ki te complementation in Māori: An unaccusative analysis. *Te Reo* 40. 45–75.

Pine, Julian M. & Lieven, Elena V.M. 1997. Slot and frame patterns in the development of the determiner category. *Applied Psycholinguistics* 18. 123–138.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry* 20(3). 365–424.

Siskind, Jeffrey Mark. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61(1–2). 39–91.

Takac, Martin & Knott, Alistair. 2016a. Working memory encoding of events and their participants: A neural network model with applications in sensorimotor processing and sentence generation. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (CogSci), 2345–2350.

Takac, Martin & Knott, Alistair. 2016b. Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: a neural network model. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (CogSci), 532–537.

Takac, Martin & Benuskova, L & Knott, Alistair. 2012. Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition* 125. 288–308.

Alistair Knott
University of Otago
alik@cs.otago.ac.nz

Martin Takac
University of Otago
takac@cs.otago.ac.nz