Using Lexical Knowledge of Verbs in Language-to-Vision Applications

Minhua Ma and Paul Mc Kevitt

School of Computing & Intelligent Systems
Faculty of Engineering
University of Ulster, Magee
Derry/Londonderry, BT48 7JL, Northern Ireland
{m.ma, p.mckevitt}@ulster.ac.uk

Abstract. In natural languages the default specification of arguments of verbs is often omitted in the surface form. The value of these arguments can be filled by lexical knowledge or commonsense knowledge of human readers, but it is a difficult task for computer programs. Here, we address the need for commonsense knowledge in computational lexicons, and discuss the requisite lexical knowledge of computational lexicons in the language-to-vision application CONFUCIUS. The underspecification problem in natural language visualisation is examined. We compare existing computational lexicons such as WordNet, FrameNet, LCS database, and VerbNet, and show how lexical knowledge in a generative lexicon can be used for disambiguation and commonsense inferencing to fill unspecified argument structures for the task of language visualisation. The possibility of lexical inference with WordNet is explored in order to extract default and shadow arguments of verbs, and in particular, the default argument of implicit instruments/themes of action verbs, which can be used to improve CONFUCIUS' automated language-to-vision conversion through semantic understanding of the text, and to make animation generation more robust by employing the commonsense knowledge included in (or inferred from) lexical entries.

1 Introduction

Underspecification in language is accepted by native speakers in everyday use. Much of what native speakers do with language is an almost unconscious procedure of retrieving and using their commonsense knowledge. However, it is a complex problem for language-to-vision applications. For example, the specification of the number and types of participants in an event expressed by a verb is crucial for a satisfactory description of its meaning. It is requisite to mold the commonsense knowledge into explicit symbolic representation structures which allow for effective content processing in language visualisation.

Human beings are visual animals and the semantics of things and their changes in the world is intimately linked to vision. Most verbs (events/states) concerning human physical activities or objects' visual properties (shape, position, colour, motion, etc.) are visually presentable. Language visualisation requires knowledge about visual as-

pects of elements in the world around us (visual semantics). Here we suggest that visual semantic knowledge about events/states be stored in a lexicon for visualisation. Furthermore, we propose that the organisation of visual semantic knowledge parallel the organisation of the generative lexicon [15], and lexical (commonsense) knowledge of verbs such as default/shadow arguments can be extracted from WordNet [6].

The long-term objective of our research is to create an intelligent multimedia storytelling system called CONFUCIUS which is capable of converting natural language sentences into 3D animation and speech. The major functionality of CONFUCIUS is automatic language-to-vision conversion. In order for such a language-to-vision conversion to be successful, systems will need to be able to exercise some commonsense reasoning and have a basic awareness of the everyday world in which they operate. A computational lexicon which contains enough lexical knowledge for commonsense reasoning is essential.

We first introduce the background of our work, the intelligent storytelling system—CONFUCIUS and review previous language-to-vision applications in section 2. Next, in section 3 we analyse existing computation lexicons WordNet, FrameNet, LCS database, and VerbNet, compare them in terms of the commonsense knowledge they contain, and suggest that generative lexicon theory is adequate for representing commonsense knowledge for filling underspecified roles. We also propose a selection algorithm based on WordNet to search default instruments/themes of verbs. Then we compare our proposal to related work in previous language-to-vision applications in section 4, and finally, we conclude and discuss further work in section 5.

2 Background and Previous Work

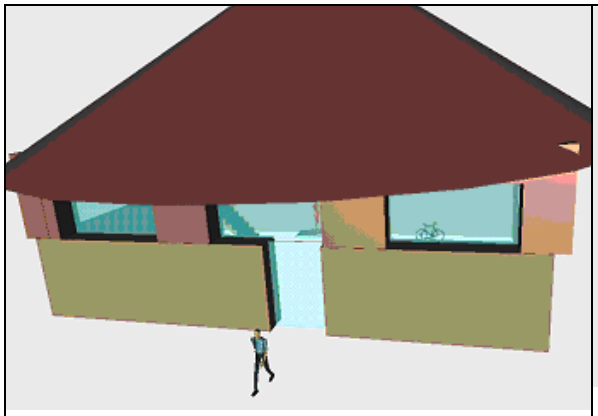
2.1 CONFUCIUS

Semantic analysis within CONFUCIUS' natural language processing module uses WordNet [6] and the LCS database [5] to fill underspecified arguments of verbs, and to perform semantic inference, disambiguation and coreference resolution. The current prototype of CONFUCIUS visualises single sentences which contain action verbs with *visual valency* of up to three, e.g. "John left the gym", "Nancy gave John a loaf of bread" [12]. Figure 1 shows examples of the 3D animation output of these single sentences.

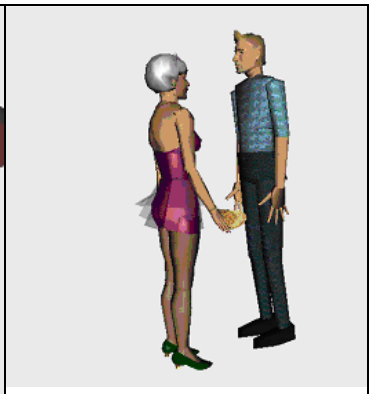
2.2 Previous Language-to-Vision Applications

There are a number of language-to-vision applications, the virtual human *Jack* [1], Wordseye [4], SONAS [8], Narayanan's iconic language visualisation [14], and their approaches to lexical/commonsense knowledge storage and retrieval vary according to different application domains. The virtual human *Jack* [1] uses knowledge from a small set of action verbs in the technical instructions domain, which is represented in a Parameterized Action Representation (PAR). Wordseye [4] obtains hypernym and hyponym semantic relations from WordNet, and has its own transduction rules and object vocabulary for commonsense inference. SONAS [8] allows the user to navigate

and interact with a 3D model of a virtual town through natural language, and it requires little lexical knowledge. Narayanan's iconic animation [14] uses Schank's scripts and conceptual dependency theory to represent and store lexical knowledge of actions. Few or none of these applications retrieve commonsense knowledge from a comprehensive computational lexicon.



a. Key frame of "John left the gym"



b. Key frame of "Nancy gave John a loaf of bread"

Fig. 1. CONFUCIUS' 3D animation output

3 Computational Lexicons

In the last decade, there have been advances on lexical knowledge of how to create, represent, organise, categorise, and access large computational lexicons, such as WordNet [6], FrameNet [3], LCS database [5], and VerbNet [10], especially for verbs, and the relation between the syntactic realisation of a verb's arguments and its meaning has been extensively studied in Levin [11].

3.1 WordNet

WordNet [6] is an electronic dictionary and thesaurus modelling the lexical knowledge of English. WordNet divides the lexicon into five categories: nouns, verbs, adjectives, adverbs, and function words. Lexical information is organised in terms of semantic relations between words. The relations used in WordNet include synonymy, autonymy, hyperonymy, hyponymy, holonymy, meronymy, troponymy (entailment), cause, value_of, attributes (has_value), and derivationally related form.

One limitation of WordNet is that it has neither predicate argument structure, nor explicit constitutive and functionality information which are important for commonsense reasoning in language-to-vision applications. This is why we have to use LCS databases to enhance the semantic analysis of CONFUCIUS' NLP module. However,

relations in WordNet do present additional semantic information implicitly. For example: the hypernym tree of "lancet" in Figure 2 contains (1) domain, (2) constitutive, (3) purpose, and (4) agentive[1] information. This feature makes lexical inference with WordNet possible, i.e. default arguments may be extracted by inference programs.

```
lancet
=> surgical knife (1)
 => knife
  => edge tool (2)
   => cutter, cutlery, cutting tool
    => cutting implement (3)
     => tool
      => implement
       =>instrumentality, instrumentation
        => artifact, artifact (4)
         => object, physical object
          => entity
```

Fig. 2. The hypernym tree of "lancet" in WordNet

Extracting Default Arguments of Verbs. Language visualisation requires lexical/common sense knowledge such as default instruments (or themes) of action verbs, functional information and usage of nouns. In Table 1, the default instruments (or themes) are the highest nodes of the hypernymy (is_a) tree in WordNet, all of whose children are possible instruments. We start from a *default assignment* [13] (of instrument/theme in this case), then propagate upward, and check if all the hyponyms of this lexical item are acceptable. If this is the case, we continue the propagation until we reach a level at which at least one of whose hyponyms is not acceptable.

Table 1. Default instruments of verbs

Verb	*Default instrument/theme (highest node of possible candidates in WordNet)*	*Example instrument/ theme*
cut	cutting implement	knife, scissors, lancet
bake	oven	oven
fry	pan	frying pan
boil	pot	kettle, caldron
drive	self-propelled vehicle	car, tractor, van
write	writing implement	pen, chalk
adorn	decoration, ornament, ornamentation	flower, jewel
kill	weapon	gun, bomb, bow

For example, "knife" is a possible instrument for the verb "cut". We propagate its hypernym tree in WordNet (Figure 3), and find that all hyponyms of "edge tool" are acceptable instruments of cutting, same for "cutter, cutlery, cutting tool", and "cutting implement". But when we propagate one more level, a quick check shows that not all children of "tool" are appropriate to serve as instruments of cutting. Some hypernyms

[1] According to Pustejovsky's [14] definition, agentive information involves the origin or *bringing about* of an object, e.g. creator, artifact, natural kind, causal chain.

of “tool”, e.g. “drill”, “comb”, cannot be used for cutting. Therefore the “cutting implement” is the highest node of possible instruments for the verb “cut” and should be stored as a default argument in its lexical entry.

```
knife
 => edge tool
  => cutter, cutlery, cutting tool
   => cutting implement
    => tool
     => implement
      => instrumentality, instrumentation
       => artifact, artefact
        => ...
```

Fig. 3. The hypernym tree of “knife” in WordNet

This approach provides a flexible specification of lexical knowledge with a proper grain size, avoiding too-particular specifications. Consider the examples (1-3), if we store “knife” as the default instrument in *cut*’s entry, it might not be appropriate for (3), whereas “cutting implement” suits all the cases.

1) John cut the bread. (bread knife)
2) The doctor cut the cancer from some healthy tissue around it. (lancet)
3) John cut the crop. (scythe)

This selection algorithm could be automated based on corpus data and linguistic ontologies. A generative lexicon [15] with this knowledge provides the capability of visualising various activities without hardcoding them as part of the animation library.

However, there is a possibility that some acceptable candidates of default instrument/theme might not have the highest nodes of possible instruments/themes in their hypernym trees. The verb “adorn” in Table 1, for instance, has “flower” as one of its possible instruments/themes. However, we cannot find the highest node of possible candidates “decoration, ornament, ornamentation” in the “flower” hypernym tree, whereas it can be found in the hypernym tree of “jewel” or “flower arrangement” (Figure 4). The need to start searching from an appropriate candidate increases the complexity of searching for default arguments in WordNet.

```
flower
 => angiosperm, flowering plant
  => spermatophyte, phanerogam, seed plant
   => vascular plant, tracheophyte
    => plant, flora, plant life
     => organism, being
      => living thing, animate thing
       => object, physical object
        => entity

jewel
 => jewelry, jewellery
  => adornment
   => decoration, ornament, ornamentation
    => artifact, artefact
     => object, physical object
      => entity
```

```
flower arrangement
 => decoration, ornament, ornamentation
  => artifact, artefact
   => object, physical object
    => entity
```

Fig. 4. Hypernym trees of "flower", "jewel", and "flower arrangement"

3.2 FrameNet

FrameNet [3] is a corpus-based computational lexicon based on the British National Corpus (BNC). It contains descriptions of the semantic frames underlying the meanings of words and the representation of the valences of words in which the semantic portion makes use of frame semantics.

Unlike WordNet which provides a framework to organise all of the concepts we use to describe the world, aiming to cover every possible subject area with at least a low level of detail, the semantic domains covered in FrameNet are limited: health care, chance, perception, communication, transaction, time, space, body, motion, life stages, social context, emotion and cognition.

FrameNet is somehow similar to efforts to describe the argument structures of lexical items in terms of *case roles* or *theta roles*, but the definition of *frame* in FrameNet is different from others, to wit, FrameNet's frames are rather semantic categories. In FrameNet, the role names[2], called Frame Elements (FEs), are local to particular conceptual structures (called frames in FrameNet); some FEs are general, while others are specific to a small family of lexical items, for instance, the *motion* frame has theme, path, source, goal, area FEs, the *activity* frame has the agent FE, whereas the *experience* frame has experiencer and content FEs.

Default arguments and shadow arguments such as instrument, means, purpose, etc. are *peripheral* FEs, and not specified in FrameNet, e.g. there are three word senses of "drive" in FrameNet's semantic domains as listed in the Table 2. The verb "drive" implies that the value of the argument instrument/means is a hyponym of vehicle. Although the knowledge is listed as a FE in the frame *operate_vehicle*, it is hard to access this information since a way to distinguish between these three frames is not provided.

Table 2. Frames and FEs of *drive* in FrameNet

Entry	*Frame*	*FEs*
drive.v.	Operate_vehicle	Area, Driver, Path Goal, Source, *Vehicle*
drive.v.	Self_motion	Area, Goal, Source, Path, Self_mover
drive.v.	Carrying	Agent, Area, Carrier Path, Theme, Path_end, Path_start

[2] The term, *role name*, covers a layer in linguistic analysis, which has been known by many other names: theta-role, case role, deep grammatical function, valency role, thematic role, and semantic frame.

Therefore, FrameNet has two limitations for language-to-vision applications: (1) its semantic domains are limited, (2) default arguments are either not contained or inaccessible.

3.3 LCS Database and VerbNet

LCS database [5] and VerbNet [10] are verb lexicons. In LCS database, verbs (approximately 9000) are organised into semantic classes and each class is represented with Lexical Conceptual Structures (LCS) [7]. LCS database defines the relation between semantic classes (based on Levin's verb classes [11]) and LCS meaning components. In a typical verb entry of the LCS database shown in Figure 5, colon is the delimiter of fields, CLASS refers to Levin's verb classes, and WN_SENSE is WordNet verb sense. Besides LCS representation and variables' specification (VAR_SPEC), a verb entry also comprises PropBank [9] argument frames and theta roles. Comparing to the above comprehensive lexicons (not only verb lexicons), LCS database does contain lexical knowledge in its selectional restrictions (variables specification), e.g. the agent of cut is specified as an animate being (VAR_SPEC ((1 (animate +)))).

```
(
 :DEF_WORD "cut"
 :CLASS "21.1.c"
 :WN_SENSE (("1.5" 00894185) ("1.6" 01069335))
 :PROPBANK ("arg0 arg1 argm-LOC(in/on-up.) arg2(with)")
 :THETA_ROLES ((1 "_ag_th,mod-loc(), instr(with)"))
 :LCS (act_on loc (* thing 1) (* thing 2)
      ((* [on] 23) loc (*head*) (thing 24))
      ((* with 19) instr (*head*)(thing 20)) (cut+ingly 26))
 :VAR_SPEC ((1 (animate +)))
)
```

Fig. 5. A verb entry of *cut* in LCS database

VerbNet [10] is also a class-based verb lexicon based on Levin's classes. It has explicitly stated syntactic and semantic information. The syntactic frames for the verb classes are represented by a Lexicalised Tree Adjoining Grammar augmented with semantic predicates, which allows for a compositional interpretation. In the verb entry of *cut* shown in Figure 6, thematic roles specify the selectional restrictions for each role like the VAR_SPEC in LCS databse, e.g. [+concrete] for the instrument of cut. Some verb senses may have more specific selectional restrictions, the verb *kick* (in the verb class hit-18.1) has the following specification:

```
Instrument[+body_part OR +refl]
Instrument[+concrete]
```

It states the instrument of kicking should be either a concrete thing or a body part.

Both LCS database and VerbNet have some form of selectional restrictions which contain lexical knowledge such as default arguments. Nevertheless, these specifications are still not enough for the language visualisation task.

```
Verb Class: cut-21.1-1
WordNet Senses: cut(1 24 25 31 33)
Thematic Roles:
  Agent[+int_control]
  Instrument[+concrete]
  Patient[+body_part OR +refl]
  Patient[+concrete]
Frames:
  Basic Transitive
  "Carol cut the bread"
  Agent V Patient
  cause(Agent,E) manner(during(E),Motion,Agent) contact (during
  (E),?Instrument,Patient)degradation_material_integrity (result
  (E), Patient)
  (other frames)...
Verbs in same (sub)class: [chip, clip, cut, hack, hew, saw,
scrape, scratch, slash, snip]
```

Fig. 6. A verb entry of *cut* in VerbNet

3.4 Comparison of Lexicons

The following Table 3 presents a comparison showing features of lexical knowledge contained in above-mentioned computational lexicons. WordNet does not have enough knowledge for compositional information of verbs, default instrument and functional information, which could be complemented by LCS database and VerbNet. However, as we mentioned earlier the selection restrictions of the instrument argument in both lexicons are insufficient for language-to-vision applications. We have to look for other sources for this knowledge.

Table 3. Comparison of verb lexicons

Lexicons	*WordNet*	*FrameNet*	*LCS DB*	*VerbNet*
Semantic domains	all	limited	all	all
POS	all	all	verb	verb
Hypernymy (is_a)	+	+	+	+
Hyponymy (n.) troponymy (v.)	+	+	-	-
Metronymy constructive (n.) compositional (v.)	+ (n.) - cause (v.)	-	+ conceptual structure	+ decompose with time func
Instrument	-	-	? selection restrictions	? selection restrictions
Functional information (telic role)	-	+ used_by	n/a	n/a

3.5 Generative Lexicon

The generative lexicon presented by Pustejovsky [15] contains a considerable amount of information that is sometimes regarded as common sense knowledge. A generative lexicon has four levels of semantic representations: *argument structure*, *event structure*, *qualia structure*, and *lexical inheritance* from the global lexical structure.

The argument structure includes *true arguments* (obligatory parameters expressed as syntax), *default arguments* (parameters which are necessary for the logical well-formedness of a sentence but may not be expressed in the surface syntax), *shadow arguments* (semantic content which is not necessarily expressed in syntax and can only be expressed under specific conditions, e.g. *Mary buttered her toast *with butter*[3]), and *adjuncts*. Qualia structure represents the different modes of predication possible with a lexical item. It is made up of *formal*, *constitutive*, *telic* and *agentive* roles. Telic roles are the function of an object or aim of an activity.

The default/shadow arguments and telic roles in a generative lexicon can complement WordNet with regard to instrument and functional information (see Table 3). Previous research in language-to-vision also proves the necessity of such information in the lexicon. In PAR [1], [2], to animate a virtual human to "walk to the door and turn the handle slowly", the representation of the “handle” object lists the actions that the object can perform, which are called telic roles in the generative lexicon theory. Wordseye [4] relies on the telic roles (functional properties) of objects to make semantic interpretations, e.g. implicit instruments, as well. To visually depict the action “ride”, it looks for objects whose functional properties are compatible with the verb to find an implied instrument “bicycle”. Hence, using a generative lexicon to make inferences on given sentences is potentially useful for language-to-vision applications where it is necessary to infer as much as possible from the given sentences.

4 Relation to Other Work

Previous language-to-vision applications hard-code commonsense knowledge, which is needed for filling in missing/underspecified information when presented in visual modalities, either into the systems’ vocabulary (Jack [1], [2], Wordseye [4]), e.g. telic roles of objects and default arguments of actions, or into a structure like Schank’s scripts [14], e.g. the prop *gun* in a *robbery* script. Here we propose a methodology to extract such knowledge from existing computational lexicons such as WordNet and store it in a generative lexicon to meet the needs of explicit information required in language visualisation.

5 Conclusion

We have argued that the language-to-vision conversion relies on lexical knowledge, such as default arguments of verbs, which may not be included in existing computa-

[3] * means illegal sentence.

tional lexicons. Existing computational lexicons, WordNet, FrameNet, LCS database and VerbNet are analysed and compared. A selection algorithm based on WordNet is proposed for finding the highest hypernym of default instruments/themes. The theory of generative lexicon shows its adequacy to fill underspecified roles. Future work will address the issue of finding appropriate hyponyms from lexical knowledge and context. For example, given the default instrument "cutting implement" for a verb sense of "cut", find an appropriate hyponym ("scythe") for the sentence "John cut the crop".

References

1. Badler, N.: Virtual humans for animation, ergonomics, and simulation. In *IEEE Workshop on Non-Rigid and Articulated Motion*, Puerto Rico, June (1997).
2. Badler, N., B. Webber, M Palmer, T. Noma, M. Stone, J. Rosenzweig, S. Chopra, K. Stanley, H. Dang, R. Bindiganavale, D. Chi, J. Bourne, and B. Di Eugenio: Natural language text generation form Task networks. Technical report. University of Pennsylvania. (1997)
3. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*, Montreal, Canada (1998)
4. Coyne, B., Sproat, R.: WordsEye: An Automatic Text-to-Scene Conversion System. Computer Graphics Annual Conference, *SIGGRAPH 2001 Conference Proceedings*, Los Angeles, Aug 12-17, 487-496 (2001)
5. Dorr, B.J., Jones, D.: Acquisition of Semantic Lexicons: using word sense disambiguation to improve precision. In Evelyne Viegas (Ed.), *Breadth and Depth of Semantic Lexicons*, 79-98, Norwell, MA: Kluwer Academic Publishers (1999)
6. Fellbaum, C. (Ed.): *WordNet: An Electronic Lexical Database*, Cambridge, MA: MIT Press (1998)
7. Jackendoff, R.: *Semantic Structures*. Cambridge, MA: MIT Press (1990)
8. Kelleher, J., Doris, T., Hussain, Q., Ó Nualláin, S.: SONAS: Multimodal, Multi-user Interaction with a Modelled Environment. In *Spatial Cognition*, S. Ó Nualláin (Ed.), 171-184, Philadelphia: John Benjamins B.V. (2000)
9. Kingsbury, P., Palmer, M.: From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation* (LREC-2002), Las Palmas, Spain (2002)
10. Kipper, K., Dang, H.T., Palmer, M.: Class-Based Construction of a Verb Lexicon. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence* (AAAI-2000), Austin, TX, USA (2000)
11. Levin, B.: *English verb classes and alternations: a preliminary investigation*. Chicago, USA: The University of Chicago Press (1993)
12. Ma, M., Mc Kevitt, P.: Visual semantics and ontology of eventive verbs. In *Proceedings of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*, Su, K.-Y., Tsujii, J.-I. (eds.), 278-285, Resort Golden Palm, Sanya, China, March. (2004)
13. Minsky, M: A Framework for Representing Knowledge. In *The Psychology of Computer Vision*, P. Winston (Ed.), 211-277, New York, USA: McGraw-Hill (1975)
14. Narayanan, A., Manuel, D., Ford, L., Tallis, D., Yazdani, M.: Language Visualisation: Applications and Theoretical Foundations of a Primitive-Based Approach. In *Integration of Natural Language and Vision Processing (Volume II)*, P. Mc Kevitt (Ed.), 143-163, London, UK: Kluwer Academic Publishers (1995)
15. Pustejovsky, J.: *The Generative Lexicon*. MIT Press (1995)