Machine Translation of Motion Verbs from English into Spanish

J. Gabriel Amores\textsuperscript{a}, Juan Pablo Mora\textsuperscript{b}

\textsuperscript{a}) University of Seville
\textsuperscript{b}) University of Huelva

Abstract

In this paper we propose an account of motion verbs within JULIETTA, an English to Spanish Machine Translation prototype. The system is implemented in Prolog, and follows a transfer approach to Machine Translation (MT). From a linguistic point of view, it is based on Lexical–Functional Grammar (Kaplan & Bresnan 1982), and has as one of its goals the solution of difficult problems in English to Spanish Machine Translation. One of those problems is the translation of motion verbs. In English these verbs encode the manner (or cause) of movement, whereas Spanish encodes the path in the verb. Thus, the appropriate translation of roll out of the room from English into Spanish would not be rodar fuera de la habitación but salir de la habitación rodando. Although verbs of this type have been studied from many perspectives, they have rarely been approached in the context of MT. JULIETTA manages to translate these sentences successfully with transfer rules based on the semantic features of the verb, thus generalizing over all verbs at once, without having to resort to particular rules for each one.
Motion events in English and Spanish

The automated translation of English and Spanish sentences with verbs of motion faces the hurdle posed by the one-to-many and many-to-one correspondences between the verbs that head them. The following sentences with the English verb *roll*, used only in its motion sense are illustrative of the difference between the two languages:

(1) The ball rolled for two hours.
   *La pelota rodó durante dos horas.*
(2) The ball rolled into the house.
   *La pelota entró en la casa (rodando).*
   lit. The ball entered the house (rolling).
(3) The ball rolled out of the house.
   *La pelota salió de la casa (rodando).*
   lit. The ball exited the house (rolling).
(4) The ball rolled up the hill.
   *La pelota subió la colina (rodando).*
   lit. The ball ascended the hill (rolling).
(5) The ball rolled down the hill.
   *La pelota bajó la colina (rodando).*
   lit. The ball descended the hill (rolling).
(6) The ball rolled across the bridge.
   *La pelota cruzó el puente (rodando).*
   lit. The ball crossed the bridge (rolling).
(7) The balls rolled together.
   *Las pelotas se juntaron (rodando).*
   lit. The balls came together (rolling).
(8) The balls rolled apart.
   *Las pelotas se separaron (rodando).*
   lit. The balls separated (rolling).
(9) The ball rolled around the pole.
   *La pelota dio vueltas alrededor del poste (rodando).*
   lit. The ball revolved around the pole (rolling).
(10) The ball rolled under the chair.
    *La pelota se metió debajo de la silla (rodando).*
    lit. The ball introduced itself under the chair (rolling).
(11) The ball rolled away.
La pelota se alejó (rodando).
lit. The ball went away (rolling).
(12) The ball rolled off the table.
La pelota se salió de la mesa (rodando).
lit. The ball went off the table (rolling).
(13) I rolled the ball for two hours.
Hice rodar la pelota durante dos horas,
lit. I made the ball roll for two hours.
(14) I rolled the ball into the house.
Metí la pelota dentro de la casa (rodando).
lit. I introduced the ball into the house (rolling).
(15) I rolled the ball out of the house.
Saqué la pelota de la casa (rodando),
lit. I took the ball out of the house (rolling).

Since the manner of movement in Spanish rodando (rolling) would usually be left out, the English verb roll would correspond to at least fifteen Spanish verbs. If the translation is from Spanish into English, the problem is magnified. A single Spanish verb such as salir could correspond to hundreds of verbs in English. A few examples will suffice:

(16) Salió del garaje en coche,
lit. exited (3rd person) the garage by car.
He drove out of the garage.
(17) Salió del garaje corriendo.
lit. exited (3rd person) the garage running.
He ran out of the garage.
(18) Salió del garaje a gatas.
lit. exited (3rd person) crawling.
He crawled out of the garage.
(19) El coche salió del garaje retumbando.
lit. the car exited the garage rumbling.
The car rumbled out of the garage.

This contrast has been studied by a number of scholars (Talmy 1985; Talmy 1991; Slobin 1996; Aske 1989; Sebastian & Slobin 1994). They have come
to the conclusion that the languages of the world can be classified into two
groups: those which encode the path or trajectory of the movement inside
the verb, and those which encode it as a satellite to the verb.\(^2\) Spanish
belongs to the first group and English to the second, as the examples above
illustrate.

Spanish will tend to express in the main verb the trajectory the object
has followed in the movement. This explains why the translation of roll in
the sentences above is a path verb, when a trajectory is present in English.
Spanish will have an inventory of these verbs which will usually be the head
of directed motion sentences. Some of them are: salir (exit), entrar (enter),
dirigirse (direct oneself), cruzar (cross), subir (go up), bajar (go down), caer
(fall), juntarse (get together), separarse (come apart), atravesar (go across),
pasar (go past/by), avanzar (go forward), retroceder (go back), meter (put
in), sacar (take out), etc.

English, on the other hand, will encode in the verb the manner or the
cause of the movement. The inventory of these verbs in English is quite
large; crawl, tiptoe, bus, drive, shuffle, pad, rustle, leap, jump, pull, push,
shove, are just a few examples (cf. Levin 1993:264–269). Typically, the
verbs encoding cause will correspond to the transitive use of the verb, if
there is one (see examples 13–15).

All this will have as a consequence that it will be easier to accumulate
a number of paths in one clause in English, since these are expressed by
non–verbal satellites. This accumulation will not be natural in Spanish.
Consider the sentence:

\[(20)\] He ran down into the cellar.

(Él) bajó y entró en el sótano.
lit. (He) went down and entered the cellar.

The Spanish translation requires two clauses to keep all the information.
On the other hand, the manner of movement will have to be expressed
by an adverbial in Spanish, although quite often it is omitted altogether,
relying on extralinguistic knowledge for its recovery. In this line, studies like
(Slobin 1996) have reported that in literary translation, going from English
into Spanish, the manner of movement is kept in 51\% of the cases, and
information about trajectory in 76\%.

It should be borne in mind that these are tendencies. Thus, English has
path verbs similar to the Spanish ones (ascend, descend, enter, advance, escape, exit, cross) but most of them are borrowings from Romance languages, and are not used widely, as the unnaturalness of the literal translations in 1 to 15 attest. Spanish, on the other hand, has verbs that encode the manner of movement. But as Slobin (1996) has pointed out, they will tend to appear without trajectories, and will tend to be interpreted as atelic, i.e., as activities rather than accomplishments (Vendler 1957). Thus, a sentence like *He ran downstairs into the cellar, but did not go in.*

(21) (a) *He ran downstairs into the cellar, but did not go in.*
(b) Corrió escaleras abajo hasta el sótano, pero no entró.
lit. (He) ran downstairs as far as the cellar, but did not go in.

The only translation into Spanish that would keep all the information would require a coordination of two clauses, with two path verbs:

(c) Bajó las escaleras y entró en el sótano, (corriendo).
lit. (He) went down the stairs and entered the cellar, (running).

As usual, the manner of movement is optional, but in this case the pause is needed because otherwise the running would be understood as applying only to the second verb.

What seems clear is that on many occasions, especially with the prepositions IN, OUT and ACROSS, the manner of movement cannot be kept in the Spanish translation and a path verb has to be used obligatorily. Some authors, like (Aske 1989), have connected this fact with the telicity or atelicity of the situation described. Manner of movement verbs can be used in Spanish with atelic events, as in Corrió escaleras abajo. With telic events, a path verb has to be used: *Entró en el sótano.* Other authors, like (Talmy
1991; Slobin 1996) have introduced the notion of boundary-crossing. Thus, when a boundary is crossed, as is usually the case with into, out, and across, Spanish must use a path verb to describe the situation.

Motion constructions in English and Spanish have been related to the resultative construction. For example, Goldberg (1995:81) views resultatives as a metaphorical extension of the caused-motion construction. The following are a few examples of resultatives:

(22) He pushed the door shut.

Cerró la puerta (de un empujón).

lit. (He) closed the door (of a push).

(23) He drank himself to death.

Se mató de tanto beber.

lit. He killed himself by drinking a lot.

Slobin and Talmy consider that, as in the case of boundary crossing, resultatives exhibit a change of state. Since languages like Spanish have a restriction against expressing these changes of state as secondary predicates, they must be encoded in the main verb.

JULIETTA

This section describes JULIETTA, a Lexical–Functional Grammar–based Machine Translation prototype implemented in Prolog, and designed originally for the translation of medical abstracts from English into Spanish (Amores 1992).

JULIETTA follows a transfer approach to Machine Translation. Any transfer–based MT system contains three main modules: analysis, transfer and generation. The input sentence is analyzed by a parser, obtaining a syntactic–semantic representation which is then transferred to an equivalent representation in the target language (Spanish). The target representation is used to generate the final translation. The transfer architecture fits perfectly with the representations assigned by an LFG grammar. In LFG, every sentence is first analyzed in terms of its constituent structure (c–structure). This level contains categorial and word–order information, which is discarded by JULIETTA during analysis. A deeper, syn-
tactic representation containing functional and grammatical information
(\(f\)-structure) is then obtained after the appropriate functional equations
have been solved. The \(f\)-structure is represented as a feature–value matrix,
common to most unification–based approaches to grammar (Shieber 1986).
Transfer takes as input a source \(f\)-structure and returns as output a tar-
get \(f\)-structure. Finally, the generation component takes an \(f\)-structure as
input and returns the corresponding \(c\)-structure.

JULIETTA is currently capable of handling most structures of the En-
glish language, including coordination (CLs, VPs, NPs and PPs), relative
clauses, complementation and modification (of verbs, nouns and adjectives),
hyphenation patterns, resultative constructions, reflexivization, ergativity,
question formation, comparative constructions, phrasal verbs, passives, etc.

Translating Motion Verbs with JULIETTA

As has been pointed out above, the translation of motion verbs from En-
glish into Spanish poses many challenges, even for human translators. As
regards Machine Translation, the translation of motion verbs seems to con-
tain a generalization which should be formalized in the system. However,
the main challenge for an MT system is that the internal structure of sen-
tences with motion verbs in English and Spanish is very different. Usually,
MT systems apply translation rules of two types: lexical transfer rules and
structural transfer rules. Lexical transfer rules trigger different transla-
tions of lexical items based on contextual information. For example, the verb
\textbf{raise} translates differently into Spanish depending on the type of object:

\begin{enumerate}
\item \texttt{raise + plant-object} = \textit{cultivar}
\item \texttt{raise + animate-object} = \textit{criar}
\item \texttt{raise + 'flag'} = \textit{izar}
\item \texttt{raise + 'question'} = \textit{plantear}
\end{enumerate}

Structural transfer rules modify the internal structure of the sentence, such
as translating a passive into an active. Motion verbs pose the problem that
it is the semantic features of the verb that trigger both a lexical transfer
rule and a structural transfer rule at the same time. In effect, we need a rule
that translates a particular construction in English into its corresponding
equivalent in Spanish. This, in turn, poses the problem of identifying the limits of the construction.

An additional problem comes from the amount of information which has to be transferred from one language to the other. It has been argued above that human translators usually lose information in their translations of motion verbs, relying on background information to fill the translation gap. However, this strategy is extremely difficult and dangerous in MT since it is almost impossible to decide what information is already known by the reader and therefore could be dispensable in the translation process.

JULIETTA uses a combination of source language and contrastive information to solve the problem. Since the motion construction may appear with different patterns, the sections that follow will describe how JULIETTA handles each one, even though the strategy is similar for all cases.

Motion verbs with one Prepositional Object

Consider examples 2–6, 9 and 10 above. In all of them the verb of movement is followed by a prepositional object headed by a preposition which indicates the path. In the suggested Spanish translation, the preposition has been translated as a verb which expresses the trajectory described by the preposition. Thus, for every preposition in English there must be a corresponding verb. Finally, the original English verb has been translated as a manner adverbial. Let us see how JULIETTA proceeds step by step. First, the lexical entry for the verb roll must be coded as vtype: motion, admitting any of the prepositions listed in its pobj specification, as follows:

\[
dic(v, roll, \text{pred: roll, ggf: [subj, pobj], vtype: motion,}
\]
\[
pobj: [\text{pcase: [across, to, into, past, up, through, down,}
\]
\[
\hspace{1cm} \text{along, under, off, out_of, away_from, around]],}
\]
\[
\hspace{1cm} \text{subj: [role: theme]]}.
\]

With this information, JULIETTA will obtain the following (simplified) result for a sentence like The ball rolled into the house:

\[
>> \text{the ball rolled into the house.}
\]
\[
pred: \text{roll}
\]
Machine Translation of Motion Verbs...

tense: past
vtype: motion
ggf: [subj, obj]
pobj: pred: house
   spec: the
pcase: into
subj: pred: ball
   spec: the

The transfer rule which performs the modifications mentioned above is shown below:

\[
(P, ojb: pcase \Rightarrow Pcase, \\
vtype \Rightarrow motion) \Rightarrow \\
([P, ([P] \Rightarrow [P1|_]), true, \\
(\text{manner}(P1, \text{Manner}), \\
\text{prep\_verb\_corresp}(Pcase, \text{Pred, OtherFeats})) \\
) \\
\Rightarrow ([\text{pred: Pred, manner: [form: Manner]|OtherFeats}]).
\]

The rule may be glossed as follows: for any predicate \textbf{P}, if it is marked as \texttt{vtype: motion}, and contains a prepositional object, then obtain its translation ([P] \Rightarrow [P1|_]) and its adverbial manner form (e.g., from \textit{roll} we would first get \textit{rodar} and then \textit{rodando}). Next, the preposition head is changed into its verbal counterpart (\texttt{prep\_verb\_corresp}(Pcase, \text{Pred, OtherFeats})). The resulting f-structure will be headed by the new predicate \textbf{Pred} followed by the manner adjunct. The result of applying this rule will be:

\texttt{pred: entrar}
\texttt{vtype: motion}
\texttt{tense: past}
\texttt{manner: form: rodando}
\texttt{subj: pred: pelota}
   \texttt{spec: el}
\texttt{pobj: pred: casa}
   \texttt{spec: el}
\texttt{pcase: en}
Which, after generation, produces:

La pelota entro en la casa rodando.

*Motion verbs with one Path particle only*

Alternatively, we may find an adverbial particle expressing the path, with no terminal location to the movement. It is the case of examples 7, 8 and 11: *The balls rolled together*. The translation rule will change minimally, but the general strategy is the same: obtain the verbal counterpart of the particle and change the original verb into a manner adjunct:

\[
([P, \text{adj2:form }= \text{ Form,}
\text{ vtype }= \text{ motion}]) \Rightarrow
([\text{adj2:}_{-}],
([P] \Rightarrow [P1|_{-}]),
\text{true},
(\text{manner}(P1,\text{Manner}),
\text{prep_verb_corresp(Form,Pred,OtherFeats)}
))
\Rightarrow ([\text{pred:Pred,adj2:[form:Manner]}|\text{OtherFeats}]).
\]

*More than one path*

As mentioned above, English permits the accumulation of paths in the motion construction. Accumulation may be achieved in two ways. First, we could add more path particles (one or two at most, according to data) before the prepositional object, as example (20) above and the following one show:

(24) He ran back down into the cellar.

\[\text{Volvíó, bajó y entró en el sótano, corriendo.}\]

lit. he came back, came down and entered the cellar, running.

In this situation, the translation rule is quite complex, since a single clause must be translated as a coordination of verb phrases in Spanish, with the manner adverbial at the end. The following rule handles this problem:
([P,pobj:adj2:form == Form,
pobj:pcase == Pcase,
 vtype == motion,
tense == T,num == N, per == Per])

>>>
([pobj:POBJ],
 ([P] => [P1]_),
 (manner(P1,Manner),
 prep_verb_corresp(Form,Pred,OtherFeats),
 lextrf(POBJ,POBJ2,POBJ),
 delfeat(adj2:_-,POBJ2,POBJ3),
 prep_verb_corresp(Pcase,Pred2,_) )

=> ([pred:Pred,
    conj:[coor:y,tense:T,num:N,per:Per,
         pred:Pred2,pobj:POBJ3,
         adj2:[form:Manner]|OtherFeats]]).

Alternatively, the event may be expanded adding more than one prepositional object with the same motion verb, as for example in:

(25) The boy ran up the hill, past the flag, into the house.
El niño subió por la colina, pasó por delante de la bandera
y entró en la casa, corriendo,
lit. The boy ascended the hill, passed the flag and entered
the house, running.

This example is even more difficult, since in addition to translating a prepositional object as a coordination of verb phrases, the original prepositional objects must also be translated and be made dependent on each of the new predicates. The rule below handles this case.

([P,pobj:pcase == Pcase,
pobj:conj:pcase == Pcase2,
pobj:conj:conj:pcase == Pcase3,
 vtype == motion,
tense == T, num == N, per == Per] ) >>>>
( [pobj:POBJ],
([P] => [P1|_]),
  ( manner(P1,Manner),
    prep_verb_corresp(Pcase,Pred1,_),
    prep_verb_corresp(Pcase2,Pred2,__),
    prep_verb_corresp(Pcase3,Pred3,__),
    leextrf(POBJ,POBJ2,POBJ),
    member(conj:Conj1,POBJ),
    member(conj:Conj2,Conj1),
    delfeat(conj:Conj1,POBJ2,POBJ3),
    delfeat(conj:Conj2,Conj1,Conj3),
    delfeat(coor:_,Conj2,Conj22),
    delfeat(coor:_,Conj3,Conj33)
  )
=> ([pred:Pred1,
    pobj:POBJ3,
    conj:[coor:comma,pred:Pred2,tense:T,num:N,
      per:Per,pobj:Conj33,
      conj:[coor:y,pred:Pred3,tense:T,num:N,
        per:Per,pobj:Conj22,
        adj2:[form:Manner]]])).

**Motion Verbs with causative meaning**

So far we have provided examples of intransitive uses of the motion verb. These verbs may also be found in transitive patterns, exhibiting the same type of transformation when they are translated into Spanish. The only difference lies in the verb associated with the preposition. If, for example, the predicate associated with the preposition *into* was *entrar*, in the causative meaning, it will be *meter*. Also, in addition to being marked as *vtype:motion*, the verb must be in its transitive use and marked *caus:yes*, for causative.

([P,pobj:pcase == Pcase,
  vtype == motion,
  caus == yes]) >>
([], ([P] => [P1|_]),true,
  ( manner(P1,Manner),

Finally, if none of the patterns above has been found, the system makes use of the default translation of the original verb.

**Results and Conclusions**

Following is a battery of examples translated with JULIETTA.

Sentence 114  the ball rolled into the house.  
La pelota entro en la casa rodando.

Sentence 115  the ball rolled out of the house.  
La pelota salio de la casa rodando.

Sentence 116  the ball rolled down the hill.  
La pelota bajo por la colina rodando.

Sentence 117  the ball rolled across the bridge.  
La pelota cruzo por el puente rodando.

Sentence 118  the balls rolled together.  
Las pelotas se juntaron rodando.

Sentence 119  the balls rolled apart.  
Las pelotas se separaron rodando.

Sentence 120  the ball rolled around the pole.  
La pelota dio la vuelta al poste rodando.

Sentence 121  the ball rolled away.  
La pelota se alejo rodando.

Sentence 122  the boy rolled the ball into the house.  
El muchacho metio la pelota en la casa rodando.
Sentence 123  the boy rolled the ball out of the house.
El muchacho saco la pelota de la casa rodando.

Sentence 124  the ball rolled for two hours.
La pelota rodo durante dos horas.

Sentence 125  he ran down into the cellar.
Bajo y entro en el sotano corriendo.

Sentence 126  he ran back down into the cellar.
Volvio, bajo y entro en el sotano corriendo.

Sentence 127  the boy ran up the hill, past the flag, into the house
El muchacho subio por la colina, paso junto a la bandera y entro en la casa corriendo.

The results show that it is possible to formalize the contrastive difference between motion verbs in English and in Spanish. A small group of transfer rules in JULIETTA is capable of translating most patterns in which motion verbs may appear in English. Nevertheless, there are some cases in which this strategy would generate the wrong translation, as for example in *He walked into the street*, which would be translated as *Entró en la calle (andando)*. Examples like these seem to be subject to pragmatic constraints, since one does not usually 'enter a street', as stated in the Spanish translation.

Notes

1. A casual examination of (Garrudo 1991, 1996) has shown that there is a large number of verbs that can be translated as *salir*. Levin (1993:234–235) lists 118 verbs of sound emission. These verbs could be translated as *salir* when followed by a prepositional phrase headed by *out*.

2. Talmy (1991:486) defines satellite as the grammatical category of any constituent other than a nominal complement that is in a sister relation to the verb root.
References


