Purcell, E.T. 1976. 'Pitch peak location and the perception of Serbo-
    Croation word tone'. *Journal of Phonetics* 4, 265-270.

Strangert, E. 1985. *Swedish speech rhythm in a cross-language perspective.*
    Stockholm: Almqvist & Wiksell.

Zetterlund, S., L. Nordstrand & O. Engstrand. 1978. 'An experiment on the
    perceptual evaluation of prosodic parameters for phrase structure
    decision in Swedish'. In E. Gårding, G. Bruce & R. Bannert (eds.),
    Nordic Prosody, 15-23. Travaux de l'Institut de linguistique de Lund
    XIII.

Choosing Aspect in Automatic Translation into Russian and Polish

Barbara Gawrońska

The paper presents an experimental procedure choosing the perfective/imperfective aspect in automatic translation from Swedish and English into typical aspect languages: Russian and Polish. The program described is based on the assumption that there are certain similarities between the (in)definiteness of Swedish/English NPs and the Slavic aspect. Both categories (aspect and definiteness) may be related to the conceptual distinction between unique referents and referents which are unmarked as to their uniqueness. The uniqueness-based approach takes into account both sentence-internal cues for aspect choice and the linguistic context of the sentence to be translated. A kind of knowledge representation is utilized as well.

Introduction

Russian and Polish – two of the five languages involved in the experimental MT-system SWETRA (Dept. of Linguistics, Lund University) are known as typical aspect languages. The lexical inventory of both Russian and Polish contains aspectually marked verb pairs, i.e. each verb (except a small group of biaspectual verbs) is inherently either perfective or imperfective. The distinction is usually marked by a prefix (Pol: *czytać/przeczytać*, R: *čitat'/ pročitat'* 'to read imp/perf') or a change in the stem, e.g. Pol: *podpisać/podpisywać*, R: *podpisat'/podpisyvat'* 'to sign perf/imp', Pol: *brać/ wziąć*, R: *brat'/vzjat'* 'to take imp/perf'. This means that a translator formulating a Polish/Russian equivalent of an English VP almost always has to choose between two members of a certain verb pair. Human translators, who are native speakers of Russian or Polish, normally perform this task without difficulty. What cues do they use when deciding which aspectual variant fits into the given context properly? Can the principles for aspect choice be formalized and used in an MT-system?

The aspect category as a linguistic problem

Do all languages express the category of aspect in some way? What exactly is expressed by this category? Questions like these have been discussed in an enormous number of works in general linguistics. Nevertheless, little agree-

ment as to the status and the meaning of the aspect category has been achieved. Some of the most common controversies in the domain of aspect-ology may be summarized as follows:

1) Shall aspect be treated as a universal category or as a language-specific one?

2) Is aspect a purely verbal category or a sentence operator, or is it pri-marily a discourse strategy?

3) Is it possible to ascribe an invariant meaning to a certain aspect value? Or must the meaning of an aspectually marked verb be derived from the se-mantic features of the verbal stem?

Each of the above questions has been answered in different ways. Several aspectologists focus on the discourse functions of aspect (Hopper & Thomp-son 1980, Wallace 1982, Paprotté 1988), others concentrate on aspect choice in isolated sentences (e.g. DeLancey 1982). There are arguments for an invariant difference between the perfective and the imperfective aspect (Forsyth 1970) as well as for investigating verbal stems one by one in order to discover the meaning of the aspect category (Apresjan 1980).

Despite all the controversies concerning the status and the main function of aspect, most researchers agree that the perfective aspect is normally chosen when referring to events, processes or states (later in the text, a more general term, viz. event-situations, will be used), which are limited, complete, countable, whereas the imperfective aspect alludes to uncompleted event-situations without clear temporal boundaries. This way of describing the distinction between the perfective and the imperfective aspect is to be found both in traditional descriptive grammars (the Soviet Academic Gram-mar 1954) and in some recent papers by cognitive grammarians (Langacker 1982, Paprotté 1988). The latter authors argue especially for a parallelism between mass names and imperfective verbs, and between countable nouns and perfective verbs. The basic conceptual distinction between spatially limited (countable) referential objects and referents without clear spatial limits (denoted by mass names) is assumed to apply to the temporal limits of event-referents: temporally bounded events become countable, i.e. perfective, and get the figure (foreground) status in a discourse, while event-situations which lack temporal limits (mass-referents) are expressed by imperfective verbs and function as discourse background.

The view on the aspect category (at least in Polish and Russian) presented in this paper is partially related to the interpretation proposed by cognitive grammarians. A similarity between typical NP-referents and event-referents is also assumed, but instead of treating the perfective/imperfective distinction as reflecting the conceptual difference between count- and mass-referents, I would prefer to relate the aspect value to another referential feature, namely, to the notion of uniqueness.

The uniqeness-based approach

The PROLOG implementation of a number of rules for aspect choice in translation from Swedish or English into Polish/Russian is based on the as-sumption that the choice between the perfective and the imperfective aspect in Russian and Polish reflects the distinction between event-situations which are marked as highly specific, unique, and those which are unmarked as to their uniqueness. By unique I roughly mean 'not identical with another refe-rent'. In Germanic languages, the referents of noun phrases may be marked as unique by the definite article or other definiteness markers, e.g. possessive pronouns. The uniqueness marking may apply both to countable and uncountable referents: *the dog is sick* refers to a specific entity of the kind dog, *the wine was good* alludes to a specific appearance of the sub-stance in question (e.g. the wine that has been drunk at a specific party). In Russian and Polish, a similar function is fulfilled by the perfective aspect – the difference being that the choice of a perfective verb marks the referent of the whole predication (an event-situation) as highly specific, unique, not identical with other event-situations named in the discourse.

The distinction between our approach and the mass/count interpretation of aspect proposed by cognitive grammarians may seem very subtle. Never-theless, it is of importance. The mass/count analogy does not account for some atypical cases of aspect use, which usually present difficulties for adult learners of Russian or Polish – e.g. the use of the imperfective aspect in Russian/Polish equivalents to a sentence like *Have you already had break-fast/lunch/dinner?* (R: *Ty uže zavtrakal/obedal/užinal?*, Pol: *Jadłeś już śnia-danie/obiad/kolację?*). The event referred to is undoubtedly finished and time-limited, i.e. countable – but in spite of these features, it is expressed imperfectively. The use of the perfective variants of the verbs exemplified is more restricted – it is e.g. possible in situations where the sender stresses the importance of the fact that a very specific food portion has disappeared,

or when a sequence of specific events is expressed – like in the example below:

R: My poobedali,   a   potom pošli   v kino
      we had-lunch-perf and later went-perf to cinema
      'We had lunch and then we went to the cinema'

Here, the perfective aspect points out that the lunch referred to was a unique one (it was followed by the action of going to the cinema), whereas in questions like:

R: Ty uže   obedal?
      you already had-dinner-imp

the sender is not interested in a unique case of having dinner, but merely in whether the addressee is hungry or not; thus, the imperfective aspect is a natural choice, although the event alluded to is a countable one.

Finding uniqueness cues

The role of the notion of uniqueness can be further illustrated by a fragment of an English text translated into Russian by a human translator. To make the example clearer, I do not quote the whole Russian text, but only specify the aspect values chosen by the translator.

Sample text 1

(the initial sentences of the preface to *An Introduction to Descriptive Linguistics* by H.A. Gleason; aspect values taken from a translation into Russian):

  1.1 Language is one of the most important and characteristic forms of human behaviour.
  (no aspect marking – a verbless predicative)

  1.2 It has, accordingly, always had a place in the academic world. (imperfective)

  1.3 In recent years, however, its position has changed greatly. (perfective)

The sample text shows that there is no clear correlation between the English tense and the Russian aspect: the aspect value may vary, although the tense value of the source text is constant (in both 1.2 and 1.3 the Present Perfect

is used). Thus, tense cannot be used as a primary cue when generating aspect. But if we look for uniqueness indices in the source text and treat them as aspect indices, the result will be quite adequate. In sentence 1.2 (*It has, accordingly, always had a place in the academic world*), the adverb *always* indicates that the predication does not refer to any unique situation – the state expressed by 1.2 may be true at any point in the time. Hence, the imperfective aspect is the only possible alternative (Polish and Russian perfective verbs in the past tense do not normally co-occur with adverbs such as *always, often*, etc). The situation expressed in 1.3 (*In recent years, however, its position has changed greatly*) contains several elements that make it contrast with the one named in 1.2. The effect of contrast is reached by the adverb *however* and by the semantics of the finite verb *changed*. In addition, the state referred to in 1.3 is placed in a quite definite time period (*in recent years*). All these factors taken together provide a sufficient motivation for marking the referent of 1.3 – in the given context – as an event-situation which is unique in relation to the generally true state mentioned in 1.2. Accordingly, the perfective aspect is used.

Sample text 1 shows that there are certain adverbials which, on their own, may be sufficient as aspect indices (e.g. *always*) and that the appropriate aspect value may be indicated by an interplay between adverbial phrases, semantic features of the main verb and the context of the current predication (as in 1.3).

An attempt to formalize some principles for aspect choice

A computer program for aspect choice in translation should take into account at least those types of aspect indices that have been observed in the sample text discussed above. The result will obviously not be a full set of aspect generating rules. However, an attempt to design an automatic procedure generating aspect is of practical and theoretical interest: the quality of translation may be improved, and an analysis of the advantages and the shortcomings of the procedure may provide a deeper insight into the nature of the aspect phenomenon.

The program presented here is implemented in LPA MacProlog and functions as an intermediate (transfer) stage in the translation process – it intervenes between the parsing of the Swedish or English text and generating of its Russian or Polish equivalents (similarily to the procedure for definiteness choice, outlined in Gawrońska 1990). For different language pairs,

slightly different variants of the transfer program are used, but all modules are based on the same main principle.

The programs used for parsing and generation are written in a modified version of Referent Grammar (Sigurd 1987), called Predicate Driven Referent Grammar (PDRG). The formalism, implemented in DCG, is an eclectic one – it incorporates features of GPSG (no transformations, use of LP-rules in parsing certain constituents, a GPSG-inspired treatment of relative clauses), LFG (c-representations and f-representations) and HPSG (the head of the phrase – especially the finite verb – playing the central role in the selection of other phrasal elements). It is just the treatment of the finite verb (or a verbless predicative) as the central element of a sentence that the name of the formalism alludes to. A PDGRG rule may be written as follows:

```
rsent(d,c_rep([advp(A),Cat1(R1,Mark1),
vfin(Verb,Aspect), Cat2(R2,Mark2)]),
f_rep([F_role1(F_Repr1),pred(Verb),
F_role2(F_Repr2),advl(Af)]),
sem_rep([event_nucl(Event),S_role1(S_Repr1)S_role2(S_Repr2),
circumstances(Feat(As))]))
--> ropadvp(A,Af,As,Feat),
rconst(Cat1(R1,F_Repr1,S_Repr1,Mark1)),
rvfin(Form,Aspect),
{rlex(Form,Verb,v,fin,Aspect,_,_,_,f_roles([F_role1(Cat1,Mark1),
pred(Verb), F_role2(Cat2,Mark2)]),
s_roles([S_role1(F_role1),S_role2(F_role2),
event_nucl(Event)]))},
rconst(Cat2(R2,F_Repr2,S_Repr2,Mark2)).

d = declarative
rsent = Russian sentence
ropadvp = Russian optional adverbial phrase
rvfin = Russian finite verb
rconst = an obligatory syntactic constituent
rlex = Russian lexical entry
```

The rule above is slightly simplified – it contains no agreement conditions and only one optional adverbial phrase. In the actual program, the number of adverbials may vary, and the subject-verb agreement is controlled.

As a result of the parsing, three kinds of representations are delivered:

1) a category representation (c_rep), which is the most language-specific one: it contains information about the surface word order, the syntactic

category of the verbal complements (the case value of the NPs, if present, or the form and the case requirement of a valency-bound preposition (this kind of information is represented by the variables Mark1 and Mark2)

2) a functional representation (f_rep), including such traditional functional roles as subject, object, predicate and adverbial

3) a semantic representation (s_rep), containing semantic roles like actor, patient, experiencer, stimulus, etc.

The rule above is a very general one: both the functional and the semantic roles (F_role1/2, S_role1/2) as well as information about their surface realizations (Cat(egory)1/2) are unspecified; in the parsing/generation process they are instantiated by utilizing the information stored in the lexical entry for the verb (the entity with the functor rlex), which may have the following shape:

```
rlex(udaril,m(hit,past),v,fin,perf,_,
agr([sg,ma]),_,
f_roles([subj(np,nom),pred(m(hit,past)),
obj(np,acc)]),
s_roles([actor(subj),patient(obj),
event_nucl(m(hit,past))])).
```

The aspect category (Aspect) is represented both in the lexical entry and in the verbal slot of the c-representation. Russian/Polish aspect is thus treated as a language specific category marked on the verb, as distinguished from the more abstract category of uniqueness, which, according to my approach, is a universal conceptual notion, expressed in different ways by different language systems.

In the translation process, the f-representation and the s-representation are utilized. After parsing an English/Swedish sentence, the program tries to find out the uniqueness value of the event expressed by the current predication using three main kinds of rules:

1) rules checking uniqueness indices inside the functional and the semantic representation without looking at the context or using knowledge representations stored in the data base

2) rules comparing the current predication with information about the prototypical predication containing the current verb (i.e. rules using a knowledge representation)

3) rules comparing the current predication with its context and inferring the probability of aspect change.

The three kind of rules apply in the above suggested order. If a rule of type 1) results in instantiating the uniqueness value of the event-referent as uni(que) or not_uni(que), the other rule types do not apply. It means that rules 1) have to discover the strongest (non)uniqueness indices – such as adverbials of indefinite frequency or duration, which normally exclude uniqueness, or other not-uniqueness indicating markers, like the English continuous tense types, aspectual verbs like *begin, stop*, etc., or Swedish constructions with coordinated verbs (such as *satt och läste*, lit. 'sat and read' – 'was reading'), which have a function similar to the English continuous tenses.

This kind of rule may be exemplified by the following, which may be used for finding habituality markers such as indefinite frequency adverbials, adverbials expressing duration or the verb *brukade* ('used to') in the Swedish input:

```
uniqueness_ind(past,sem_rep(Slist),not_uni):-
in_list(Functor(Repr,Feature)),Slist),
uniqeness_relevant(Functor),
not_uni(Tense,Functor,Feature).
```

Slist is the semantic representation (formulated as a Prolog list). The predicate in_list checks if an element is a member of the list Slist. The functor of a list member (Functor) may be defined (in the data base) as potentially relevant for the uniqueness value (uniqueness_relevant). For example, functors like frequency or duration, or action_kind are treated as uniqueness-relevant. Thus, if the semantic representation Slist contains an element like: action_kind(m(use,past),habituality), i.e. the representation of the verb *brukade*, or frequency(often,indef), i.e. the representation of the adverb *ofta* (often), then the program has to check whether the combination of the functor, the feature specified inside the brackets (e.g. indef or habituality) and the tense value (here: past) results in a specific uniqueness value. As the data base contains the following information:

```
not_uni(past,frequency,indef).
not_uni(_,_,habituality).
```

the program will decide that a sentence in the past tense containing an adverb like *ofta* or a finite verb like *brukade* does not refer to a unique event-situation. As a consequence, the imperfective aspect will be preferred when generating the target equivalent.

The next step is – given the semantic representation and the uniqueness value – to create a new functional representation, if needed, and then the appropriate c-representation. Sometimes, the input and the output may have the same f-representation, and differ only as to some details in their c-representations, like e.g. simple transitive sentences:

Sw: pojken   slog   ofta   hunden
      the-boy hit   often the-dog

Pol: chłopiec   często   bił   psa
      boy-nom   often   hit-imp   dog-acc/gen

      f_rep([subj(m(boy,sg)),pred(m(hit,past),
      advl(often,indef)]])

But in cases such as the Swedish construction with *brukade* there is a need to change the functional representation, since the most natural way of expressing the feature habituality in the Russian or Polish equivalent is by using the imperfective aspect and (optionally, if the habituality should be emphasized) an adverb like Russian *obyčno* 'usually'. Such changes are not especially difficult to implement if the semantic representation is used as a kind of interlingua. In the s-representation, the infinitive following the habituality-marking verb *brukade* is treated as a semantic kernel of the event situation. The program has therefore to find the target equivalent of the semantic kernel, and make it the main predicate, provide the target representation with the right aspect value and then – optionally – insert an adverbial as an extra habituality marker. These operations result in translations like:

Sw: Han brukade komma för sent
      he   used   come   too late

Pol: Zwykle   się   spóźniał
      usually he-was-late-imp

Rules belonging to types 2) and 3) take care of cases lacking such obvious uniqueness indices as in the example above. Type 2) has access to the descriptions of most typical predications containing a certain verb. Such descriptions are formulated (simplified) as e.g.:

```
proto_event(become_engaged,[actors([specific,limited_ref(2),]),
duration(limited),frequency(low,def),uniqueness(high)]).
```

A type 2) rule applying to a predication containing the predicate meaning 'be engaged' checks whether the actors involved are two specific individuals and whether there is any violation of the other conditions specified in the description of the prototypical event. If the current predication matchs most of the elements specified in the frame proto_event, the uniqueness value of the proto_event (here: uniqueness (high), which means: unique with a high degree of probability) will be ascribed to the current event-referent. This means that, when translating a Swedish sentence like *Per och Lisa förlovade sig* 'Per and Lisa became engaged' the perfective aspect would be chosen, whereas the same Swedish verb used in a sentence like: *Förr i tiden förlovade folk sig på föräldrarnas order* 'In former times, people got engaged by order of their parents' would be rendered by the Russian/Polish imperfective verb.

An example of an type 2) rule is the following one:

```
uniqueness_ind(past,sem_rep(Slist),not_uni):-
in_list(event_nucl(m(EventNucl_)),Slist),
proto_event(EventNucl,Condlist),
in_list(uniqueness(high),Condlist),
not(cond_matching(Slist,Condlist)).
```

This rule states simply that if the prototypical event (proto_event) containing the semantic kernel of the current predication (EventNucl(eus)) is specified as unique with a high degree of probability (expressed in Prolog by the constant: uniqueness(high)) and if the elements of the semantic representation of the current sentence do not match the conditions stored in the entity proto_event, then the uniqueness value of the event-situation referred to is not unique. Writing specific rules matching semantic representations with proto-events is obviously not a trivial task – there are few prototypical events which are as easily described as the case of being engaged.

Type 3) rules are the most complicated ones, as the task performed is to compare the current predication both with the prototypical event and with

the previously stored semantic representations (including their uniqueness values) in order to discover a possible motivation for aspect change. At the present time, only a restricted number of cues have been implemented. The program utilizes principles like:

– it is quite probable that parts of a unique event may also be unique, if no counter-indices (as e.g. indefinite duration markers) have been found
– it is probable that a predication which describes the manner of performing an already introduced event should be treated as imperfective (it expresses a property of an event-referent, and may be compared to a predicative NP, which does not introduce a new referent, but ascribes a property to an already introduced one)
– adverbials marking a kind of opposition (*however*, etc.) and their interplay with other adverbials may be important cues for aspect change.

Conclusions

The main problems when implementing a procedure for aspect generation are to formulate concise and coherent descriptions of prototypical events, to design an appropriate hierarchy of rules comparing the current predication with the prototypes and to describe conditions for aspect change. This is a field for further research. Nevertheless, some uniqueness indices are possible to formalize and to implement in an MT-system (obviously, a system accepting lexical and syntactic restrictions). The above approach is a kind of compromise between different points of view represented in current research on aspect: the overt aspect is treated as language-specific, but the conceptual distinction behind the aspect choice is assumed to be based on the universal notion of uniqueness; furthermore, both sentence-internal and contextual factors are taken into consideration. The compromise seems to be quite useful.

References

Apresjan, Ju. D. 1980. *Principy semantičeskogo opisanija edinic jazyka. I: Semantika i predstavlenie znanij*. Trudy po iskusstvennomu intellektu II. Tartu: Riikliku Ulikooli Toimetised.

DeLancey, S. 1982. 'Aspect, transitivity, and viewpoint'. In P.J. Hopper (ed.), *Tense – aspect: Between semantics & pragmatics*, 167-184. Amsterdam: John Benjamins Publishing Company.

Forsyth, J. 1970. *A grammar of aspect. Usage and meaning in the Russian verb*. Cambridge: University Press.

Gawrońska, B. 1990. 'Translation great problem – on the problem of inserting articles when translating from Russian into Swedish'. *Proceedings of COLING-90*, vol. 2, 133-138.

Grammatika russkogo jazyka (Soviet Academic Grammar), vol. 2. 1954. Moskva: Nauka

Hopper, P.J. & S. Thompson. 1980. 'Transitivity in grammar and discourse'. *Language* 56.2, 251-299.

Langacker, R. W. 1982. 'Remarks on English aspect'. In P.J. Hopper (ed.), *Tense – aspect: Between semantics & pragmatics*. 265-304. Amsterdam: John Benjamins Publishing Company.

Paprotté, W. 1988. 'A Discourse perspective on tense and aspect in Standard Modern Greek and English'. In B. Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*, 447-505. Amsterdam: John Benjamins Publishing Company

Sigurd, B. 1987. 'Referent grammar (RG). A generalized phrase structure grammar with built-in referents'. *Studia Linguistica* 1987:2, 115-135.

Wallace, S. 1982. 'Figure and ground: The Interrelationships of linguistic categories'. In P.J. Hopper (ed.), *Tense – aspect: Between semantics & pragmatics*. 201-223. Amsterdam: John Benjamins Publishing Company.

Variation and Deviation in Language Acquisition
Some Hypotheses and Preliminary Observations

Gisela Håkansson[1], Ulrika Nettelbladt[2] and Kristina Hansson[2]

A study was made regarding the acquisition of specific word order patterns in Swedish. Three groups of language learners were studied cross-sectionally, viz. children with first (L1) and second (L2) language acquisition and children with disordered language development (LD). The results revealed interesting differences, especially between L1 and L2 children in terms of their word order preferences. The L1 children used a varied word order in declaratives, whereas the L2 children kept a strict word order pattern. The results from the LD children do not give a clear-cut picture and there are great individual differences. This indicates that LD children constitute a more heterogeneous group than the L2 children.

Introduction

A comparison between grammatical development in second language learners and children with grammatical disabilities as a part of a specific, developmental language disorder[3] shows striking similarities between the two groups (Nettelbladt & Håkansson 1991). Both groups make errors of omission and overgeneralization; for example, omissions of functional morphemes and overgeneralizations of inflections and word order patterns. In spite of great potentials both theoretically and practically, there have been surprisingly few connections between the two research areas of second language acquisition (henceforth L2) and of developmental language disorders (henceforth LD). An important exception is a study by Hyltenstam and Magnusson 1981.

[1]Dept. of Linguistics.
[2]Dept. of Logopedics and Phoniatrics.
[3]According to Nettelbladt 1983, specific developmental language disorder is a diagnostic term assigned to children who do not develop their language skills in appropriate time as compared to children with normal language development. Such children are usually referred from child welfare centres. Those assessed to have a more severe disorder are given special intervention by speech and language clinicians.