A Prototype of an English-Polish Machine Translation System

Anna Kupść, Teruko Mitamura, Eric Nyberg
2002
CMU-CS-??-???

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

To appear in CMU technical reports series.

Abstract

In this paper, we discuss issues related to building a prototype of a high-quality knowledge-based English-Polish machine translation (MT) system. This system is an extension of the existing multilingual KANTOO MT system which aims at translation of domain-specific technical documentation. We address challenges for MT in generation of Polish, a language with rich inflection, morphologically marked aspect and free word order.
**Keywords:** knowledge-based, domain-specific machine translation, English-Polish machine translation, Polish, unification grammars
Contents

1 System description 3

2 Basics of Polish 6

3 Morphology 8

4 Case Assignment 9
   4.1 Genitive of Negation .................................................. 9
   4.2 Genitive of Negation in Verbal Complexes ......................... 11

5 Aspect 14
   5.1 Declarative Mood .......................................................... 14
   5.2 Imperative Mood ........................................................... 16
   5.3 Infinitives ................................................................. 17

6 Personal vs. Impersonal Forms 19

7 Middle Voice 21

8 Word Order 22
   8.1 Reversed Subject .......................................................... 22
   8.2 Sentential vs. Verbal Adjuncts ......................................... 24
     8.2.1 Verbal Adjuncts ...................................................... 25
     8.2.2 Sentential Adjuncts ................................................ 26

9 Pronominal Anaphora Resolution 27
   9.1 Subject Pronouns ......................................................... 30
   9.2 Object Pronouns .......................................................... 30

10 Comparison 32
   10.1 Case Assignment ........................................................ 32
   10.2 Aspect ................................................................. 32
   10.3 Word Order .............................................................. 33
   10.4 Anaphora Resolution ................................................... 33
   10.5 Summary Results ....................................................... 34

11 Conclusions 35

A Verbal Inflection 36

B Noun Inflection 37
   B.1 Masculine Nouns ......................................................... 37
B.2 Feminine Nouns ......................................................... 38
B.3 Neuter Nouns ......................................................... 38

C Test Suite ................................................................. 39
C.1 Case Assignment ...................................................... 39
C.1.1 Verbs ................................................................. 39
C.1.2 Prepositions ......................................................... 39
C.1.3 Nouns ................................................................. 39
C.1.4 Relative Clauses ................................................... 39
C.2 Aspect ................................................................. 39
C.2.1 Main Clause ......................................................... 39
C.2.2 Subordinate Clause .............................................. 40
C.3 Word Order .......................................................... 40
C.3.1 Reversed Subject ............................................... 40
C.3.2 Verbal Adjuncts .................................................. 40
C.3.3 Sentential Adjuncts ............................................. 40
C.4 Anaphora Resolution .............................................. 41
C.4.1 Object Pronouns ............................................... 41
C.4.2 Subject Pronouns .............................................. 41
C.4.3 Interclausal Anaphora ...................................... 41
In this paper,¹ we discuss issues related to building a prototype of a high-quality knowledge-based English-Polish machine translation (MT) system. This system is an extension of the existing multilingual KANTOO MT system which aims at translation of domain-specific technical documentation. We address challenges for MT in generation of Polish, a language with rich inflection, morphologically marked aspect and free word order.

The organization of the paper is as follows: in §1, we briefly describe general system architecture, whereas §2 provides some linguistic background on Polish. Then, in sections 3–9 we discuss specific issues related to generation of Polish, such as inflection (§3), case assignment (§4), aspect assignment (§5), personal and impersonal forms (§6), middle voice (§7), word order (§8) and anaphora resolution (§9). The two final sections, §10 and §11, contain a brief comparison with other systems and conclusions, respectively.

1 System description

The current English-Polish MT project is an extension of the existing multilingual KANTOO system (a reimplemention of the KANT system, cf. Mitamura et al. (1991), Mitamura and Nyberg (1992)) developed at Carnegie Mellon University. KANTOO is a knowledge-based, domain-specific MT system (in the English-Polish MT project, the domain is restricted to printer manuals) and it uses Interlingua (IR) as a semantic representation, see Leavitt et al. (1994). The system takes as the input a text written in constrained English (controlled language), which limits vocabulary and grammar of sentences accepted by the system, cf. Kamprath et al. (1998). Example (1) presents a sample English input along with the IR representation and its Polish translation provided by the system.

(1) The printer prints pages.

(*A-PRINT
 (agent
  (*O-PRINTER
   (number singular)
   (reference definite)))
  (argument-class agent+theme)
  (mood declarative)
  (punctuation period)
  (tense present)
  (theme
   (*O-PAGE

¹Research reported in this paper has been partly supported by the post-doctoral NATO scholarship held by Anna Kupść (Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland) during her stay at Carnegie Mellon University in October 2001 – April 2002. We wish to thank Krzysztof Czula for his comments and detailed discussions of various aspects of the work presented here, as well as Kathy Baker, Annelen Brunner, Adam Przępórkowski, Agata Savary, Małgorzata Styś-Budzikowska, Enrique Torrejon, Magdalena Wóblek for their comments on earlier drafts of this paper. We are also grateful to Adam and Danuta Przępórkowsky, Małgorzata Marciniak, Agnieszka Mykowiecka and Elżbieta Hajnecz for their native-speaker judgements and to David Svoboda and Michael Duggan for technical support.
Drukarka drukuje strony.

The Polish generator module consists of four components: a mapper, a unification grammar (a type of context-free grammar), a morphological generator and a post-processing module. The general architecture of the generator module is presented in fig. 1.

![Diagram of the generator architecture]

**Figure 1: KANTOO Generator Architecture**

Mapping rules transform the IR semantic representation of the source text into a syntactic structure corresponding to the output in the target language. The structure is a functional structure or FS in the LFG (Lexical Functional Grammar) formalism, cf. Bresnan (1982). An FS for Polish, corresponding to the IR in (1), is given in (2).

(2)  
    (aspect imperf)  
    (cat v)  
    (class 2a)  
    (form fin)  
    (mood declarative)  
    (obj (  
        (agr ((gender fem) (number plural) (person 3)))  
    )  
)
(cat n)
(class f1)
(gender fem)
(root strona))
(punctuation period)
(root drukuje)
(subj
  (agr ((gender fem) (number singular) (person 3)))
  (cat n)
  (class f1)
  (gender fem)
  (root drukarka))
(tense present))

Generation grammar rules convert this FS into a list of lexical tokens (FS frames), (3), which are then fed to the morphology module responsible for generating appropriate inflected forms.

(3)  [(cat x) (root @cap)],
    (case n)
    (cat n)
    (class f1)
    (gender fem)
    (number singular)
    (person 3)
    (root drukarka),

    (aspect imperf)
    (cat v)
    (class 2a)
    (form fin)
    (gender fem)
    (mood declarative)
    (number singular)
    (person 3)
    (root drukuje)
    (tense present)),

    (case a)
    (cat n)
    (class f1)
    (gender fem)
    (number plural)
    (person 3)
    (root strona)},

5
((cat x) (previous-space -) (root period))]

Finally, a set of post-processing rules is applied to produce the resulting surface form of translation, as in (1), by cleaning up spacing, e.g., (previous-space -), capitalization, e.g., (root @cap), punctuation, e.g., (root period) etc.

In this paper, we will be mostly concerned with three generation modules: mapper, grammar and morphology. In order to develop the current prototype, a small corpus of about 280 English sentences from a printer manual has been examined. Linguistic phenomena and vocabulary of this corpus served as the basis for the system prototype.

2 Basics of Polish

Polish is a West Slavic language. Like many other Slavic languages, Polish is an inflectional pro-drop language with a relatively free word order; additionally there are no definite / indefinite articles in Polish. The relevant aspects of the language are presented below.

Polish nouns are inflected for case and number and have a grammatical gender. Traditionally, three genders are distinguished in Polish: masculine, feminine and neuter. However, due to various agreement types a more fine-grained classification is often necessary, cf. Mańczak (1956), Saloni (1976), Czuba (1997). For the purposes of the prototype described in this paper, the following genders are distinguished: masculine-human (e.g., odbiorca ‘recipient’), masculine (non-human, e.g., komputer ‘computer’), feminine (e.g., strona ‘page’) and neuter (e.g., gniazdko ‘jack’).

Polish has seven morphological cases: nominative (nom), genitive (gen), dative (dat), accusative (acc), instrumental (inst), locative (loc) and vocative (voc). An example of inflection paradigm for the feminine noun drukarka ‘printer’ is given in (4).

<table>
<thead>
<tr>
<th>CASE</th>
<th>SINGULAR</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom</td>
<td>drukarka</td>
<td>drukarki</td>
</tr>
<tr>
<td>gen</td>
<td>drukarki</td>
<td>drukarek</td>
</tr>
<tr>
<td>dat</td>
<td>drukarce</td>
<td>drukarkom</td>
</tr>
<tr>
<td>acc</td>
<td>drukarkę</td>
<td>drukarki</td>
</tr>
<tr>
<td>inst</td>
<td>drukarką</td>
<td>drukarkami</td>
</tr>
<tr>
<td>loc</td>
<td>drukarce</td>
<td>drukarkach</td>
</tr>
<tr>
<td>voc</td>
<td>drukarko!</td>
<td>drukarki!</td>
</tr>
</tbody>
</table>

Since use of vocative in technical documentation is very limited, for simplicity, this case is not taken into account in the system.

Polish verbs are inflected for person, number, gender, tense and mood while aspect is a derivational category. Polish verbs may have two aspect forms: imperfective, e.g., drukuje ‘prints’, or perfective, e.g., wydrukuję ‘will print_{3sg} (out)’. Aspect is independent of tense as it is also present on infinitives: drukować ‘to print_{imperf}’ and wydrukować ‘to print_{perf}’. Perfective verbs have no present tense and are used as non-analytical future tense forms. As far as mood is concerned, Polish has declarative, conditional and imperative moods. A
part of inflectional paradigm for the aspectual pair *drukować* ‘print$_{imperf}$’ vs. *wydrukować* ‘print$_{perf}$’ is given below:

\[
\begin{array}{|c|c|c|c|}
\hline
\text{PERSON} & \text{masc-hum&masc} & \text{fem} & \text{neut} \\
\hline
1\text{st} & \text{drukowałem} & \text{drukowałam} & \text{drukowałem} \\
2\text{nd} & \text{drukowałeś} & \text{drukowałaś} & \text{drukowałoś} \\
3\text{rd} & \text{drukował} & \text{drukowała} & \text{drukowało} \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{PERSON} & \text{masc-hum&masc} & \text{fem} & \text{neut} \\
\hline
1\text{st} & \text{wydrukowałem} & \text{wydrukowałam} & \text{wydrukowałem} \\
2\text{nd} & \text{wydrukowałeś} & \text{wydrukowałaś} & \text{wydrukowałoś} \\
3\text{rd} & \text{wydrukował} & \text{wydrukowała} & \text{wydrukowało} \\
\hline
\end{array}
\]

Polish has also a number of participles, e.g., present adverbial participle (for imperfetive verbs only, e.g., *drukując* ‘while printing’), past adverbial participle (for perfective verbs only, e.g., *wydrukowawszy* ‘having printed’), active adjectival participle (for imperfective verbs only, e.g., *drużjący* ‘printing’), passive adjectival participle (perfective: *wydrukowany* ‘printed’, imperfective: *drukowany* ‘being printed’), impersonal participle (perfective: *wydrukowano* ‘(one) printed’, imperfective: *drukowano* ‘(one) was printing’) and a gerund nominal form (perfective: *wydrukowanie* ‘printing (out)’, imperfective: *drukowanie* ‘printing’).

Adjectives in Polish are inflected for case, number and gender. An example of inflectional paradigm for the adjective *dobry* ‘good’ is given in (7).
Adjectives and adverbs are gradable but in the current system only the positive degree is assumed. Prepositions and adverbs are uninflected in Polish. Other features of Polish will be discussed, if necessary, in the rest of the paper.

3 Morphology

The Polish morphology module developed in the system implements inflection of nouns, verbs, adjectives as well as relative, demonstrative and personal pronouns. Additionally, derivation of contemporary adverbial participles, passive adjectival participles and gerunds has been incorporated.² The current prototype implements only inflectional classes and groups necessary for translation of the test corpus. Moreover, due to complexity of Polish inflection, some less frequent stem variations have not been implemented but the corresponding inflected forms have been listed as irregular forms.

For verbs, classification of verb groups comes from Tao (1990), as implemented in Czuba (1995), see Appendix A, and is based on the third person singular present tense form. This form is considered the base for verb inflection, which allows us to provide a more systematic classification and simplify stem alternations. In the system, the base for inflection is indicated by the root attribute. We retain this name as the attribute is hard-wired in the system but in the present context this name is a misnomer.

For the current prototype, 14 out of 22 inflectional verb groups have been implemented (groups 1, 2a, 2b, 2c, 3c, 4a, 4b, 5a, 6a, 7a, 7b, 9a, 10a and 10b). The following verb forms can be generated: infinitive, imperative (2nd person singular), present and past tense personal (second singular and third person forms; no first person forms and second person plural), including an overt gender distinction in past tense forms. As for future tense forms, only non-analytical forms (i.e., perfective verbs) are used. Non-past inflection of perfective verbs coincides with that of present tense imperfective verbs. Therefore, in order to reduce the number of morphological rules, perfective non-past forms are assigned present (rather than future) tense in the system.

Inflection of nouns is based on the classification proposed in McShane (2001), see Appendix B. For masculine and human-masculine nouns, the singular genitive form of a noun is considered the base for inflection. For other nouns, the singular nominative form is taken as the base.

²Due to time limitation of the project, inflection of numerals has not been encoded.
7 masculine nominal groups (1, 2, 3, 5, 6, 7, 8) and 1 masculine-human group (17) have been incorporated. In order to avoid redundancy, groups with similar inflectional patterns (1 and 3; 5 and 7; 6 and 8) have been merged. For feminine nouns, groups 1, 3, 5, 7, 9 and 12 (6 out of 15) have been implemented (again, similar groups 7 and 9 have been merged). Finally, for neuter nouns, groups 1, 2 and 4 have been incorporated (3 out of 7). Additionally, one neuter nominal group has been added for uninflected, usually foreign, neuter nouns such as menu or polo. Although gerunds are derived from verbs (unlike other nouns, their base form is that of the corresponding verb), morphologically they are neuter nominals. Therefore, paradigm of neuter nouns has been appropriately extended to incorporate inflection of gerunds.

Polish adjectives have a very regular inflectional pattern and no detailed classification was necessary. The adjectival paradigm includes case, number and gender inflection whereas only positive degree is considered. Passive adjectival participles are derived from verbs but they have adjectival inflection. In the current system derived forms are used only in nominative and, hence, only these forms are available (no other case forms have been implemented).

The relative pronoun który ‘which’ has adjectival inflection and satisfies the set of morphological rules for adjectives. For other pronouns (demonstrative and personal), separate rules have been implemented.

Finally, adverbial contemporary participle is derived from imperfective verbs. Just like adverbs, however, it is uninflected (only positive degree forms are implemented).

4 Case Assignment

Unlike English, Polish is not a positional language and grammatical categories are marked by case rather than by their position in a sentence. Hence, case assignment plays an important role in expressing the meaning of a sentence.

4.1 Genitive of Negation

Case assignment is often considered a lexical phenomenon and encoded in lexical entries of predicates which assign case to their arguments. Such an approach, however, cannot be successfully applied to Polish. As (8) indicates, case assignment in Polish is sensitive to negation.

(8) a. The printer prints a page.
   Drukarka drukuje stronę.
   printer_{nom} prints  page_{acc}

   b. The printer does not print a page.
   Drukarka nie drukuje strony / *strony.
   printer_{nom} not prints  page_{gen} / page_{acc}

3We concentrate here on case assignment to NP complements of verbal predicates. In the system, case value of NP complements of prepositions is specified at the lexical level, whereas complements of nouns are assigned case in the grammar.
In (8), an accusative (acc) complement of a non-negated verb has to appear in genitive (gen) if the verb is negated. This phenomenon is called genitive of negation (GoN) and is fairly regular in Polish. If the accusative complement of drukuje ‘prints’ in (8a) was assigned case in the lexicon, the negated verb would be expected to appear in accusative as well, which is ungrammatical, as indicated in (8b).  

On the other hand, complements of some verbs, e.g., zarządzac ‘to operate’, are immune to negation.

(9)   a. This software operates the system.
      Oprogramowanie zarządu systemem.
      software_{nem} operates system_{inst}

   b. This software does not operate the system.
      Oprogramowanie nie zarządu systemem.
      software_{nem} not operates system_{inst}

In order to account for the contrast between (8a) and (8b), on the one hand, and between (8) and (9), on the other, we adopt case assignment analysis of Przepiórkowski (1999), based on a cross-classification of Polish morphological cases into lexical vs. structural (similar to inherent vs. structural case dichotomy in Chomsky (1986)). The former are immune to the syntactic environment and are resolved in the lexicon whereas the latter are context-sensitive and are assigned in syntax.

According to the data presented above, instrumental (inst) is a lexical case: despite negation, the complement of zarządzac ‘operates’ remains unchanged in (9). On the other hand, the case of the complement in (8) depends on negation and accusative (acc) and genitive (gen) are structural cases. They are assigned by syntactic rules and unspecified in the lexicon.

This idea is implemented in the system by using FS features: obj, which represents NP complements, and neg, which indicates verbal negation. For verbs which assign lexical case to their NP complements, e.g., zarządzac ‘operates’ in (9), the obj case value is specified in the lexicon.

(10) zarzadza/c-v = (*CR* ((morph verb-imperf) (root zarzadza) (class i)
                          (obj ((case i))))
               ((morph verb-perf) (root zarzadzi) (class 10b)
                          (obj ((case i))))),

On the other hand, for verbs which require structural case on their NP complements, e.g., drukuje ‘prints’ in (8), the obj case value is not specified in the lexicon, (11), but is assigned in the grammar by syntactic rules.

(11) drukowa/c-v = (*CR* ((morph verb-imperf) (root drukuje) (class 2a)
                          ((morph verb-perf) (root wydrukuje) (class 2a))),

A syntactic rule which takes into account case variation triggered by negation is presented below.

---

4 This argumentation is based on the assumption that Polish verbal negation is not a lexical phenomenon.
(12)  (vp >-> (v <np>)
     ((%x0 cat) =c (*OR* v advp)) ;; cat value check
     (%x0 obj) = *DEFINED*) ;; check if obj is defined
     (%x2 = %x0 obj)) ;; np becomes obj
     (*TRY* ((%x2 case) = *UNDEFINED*)) ;; if obj’s case is undef
     (*TRY* (%x0 neg) =c +) ;; if the verb is negated
     (%x2 case) = g)) ;; assign gen
     ((%x2 case) = a))))) ;; else, assign acc
     (%x1 = %x0)) ) ;; v assigned other vp features

The *TRY* statements do the actual case verification and assignment. It is first checked
whether the complement is assigned case, i.e., whether the value of the case path is defined.
If it is not, verbal negation is checked next. The complement takes genitive if the verb is
negated and accusative otherwise. If the case is already assigned, i.e., it has been specified
in the lexicon, the rule in (12) does not change it. *TRY* is successfully executed only if one
of the tried statements is true. Otherwise, the rest of the code is executed.

4.2 Genitive of Negation in Verbal Complexes

Case assignment is in general a local phenomenon, i.e., an argument is assigned case by its
governing category. In Polish verbal complexes (VC), however, negation of a higher verb
can influence the case value of an argument of a lower (non-negated) verb. Let us consider
the following examples.

(13)  a. You can scan a document.
     Można zeskanaować dokument.
     canimpers scaninf documentacc
     b. You cannot scan a document.
     Nie można zeskanaować dokumentu.
     not canimpers scaninf documentgen

(14)  a. Try sending a fax.
     Spróbuj wysłać faks.
     try sendinf faxacc
     b. Do not try sending a fax.
     Nie próbuj wysłać faksu.
     not try sendinf faxacc

(15)  a. Try operating the system.
     Spróbuj zarządzać systemem.
     try operateinf systeminst
     b. Do not try operating the system.

A verbal complex (VC) is a sequence of verbs such that one subcategorizes for another without an
intervening subordinate conjunction such as że, żeby ‘that’. In (13), the verb *można* ‘can’ and its infinitival
complement zeskanaować ‘to scan’ form a VC.
Nie próbuj zarządzać systemem.

As before, an instrumental complement is immune to negation, (15), whereas accusative changes to genitive in (13) and (14). In IR, modal verbs are represented by a set of binary features whereas their complement is considered the main predicate; IR for (13b) is given in (16).

(16)  (*A-SCAN
   (ability +)
   (agent
     (*G-PRONOUN
      (number
        (:OR plural singular))
      (person second)
      (reference definite))
     (argument-class agent+patient)
     (mood declarative)
     (negation +)
     (patient
      (*O-DOCUMENT
       (number singular)
       (reference definite)))
     (punctuation period)
     (tense present))

Therefore, in the IR representation of (13b) (and also in the FS), negation is in fact present on the verb whose argument is influenced by GoN. Hence, the previously introduced case assignment principle is able to resolve case value in (13b).\(^5\)

The verb ‘try’ in (14), however, is not a modal and it is the main predicate in IR. Hence, negation is not directly available for the verb wystąć ‘to send’ whose complement is affected; IR for (14b) is given in (17).

(17)  (*A-TRY
   (argument-class agent+patient)
   (mood imperative)
   (negation +)
   (patient
    (*A-SEND
     (agent
      (*G-GAPPED-ARGUMENT
       (gapped +)))
     (argument-class agent+theme)

\(^5\)In fact, since different syntactic rules are used to generate VPs with an NP or a VP[inf] complement, the same case assignment principle is present in different syntactic rules.
(mood declarative)
(nominal +)
(number mass)
(reference nc-reference)
(tense present)
(theme
  (*O-FAX
   (number singular)
   (reference indefinite))))
(punctuation period)
(tense present))

In order to account for the case change, we have to transmit information about negation of the main verb to its (non-negated) infinitive complement.\footnote{We assume here a hierarchical structure of VCs, i.e., the main verb subcategorizes for a VP[inf] complement.}

In addition to neg, which indicates negation of the verb, we introduce an auxiliary feature fake-neg for its infinitive complement. The values of fake-neg and neg are equal. Hence, if the verb is negated, i.e., (neg +), its infinitive complement is considered negated too, (fake-neg +). The appropriate fragment of the grammar rule which does the actual assignment is presented in (18).

\begin{verbatim}
(18)  (<vp> ==> (<vp> <xcomp>)
       ((*TRY* ((%x0 neg) = *DEFINED*)
          (%(x2 fake-neg) = %(x0 neg)))) ;; assign fake-neg value
       ...)) ;; to the inf complement
\end{verbatim}

This modification allows GoN to be triggered by two factors: true local negation (neg) or negation on a higher verb (fake-neg). If either of them is present, a complement which is not assigned case is assigned genitive; otherwise the complement has to be accusative. If a complement has been assigned case, its case value does not change. The version of (12), which takes into account this modification, is presented below.

\begin{verbatim}
(19)  (<vp> ==> (<v> <np>)
       (%(x0 cat) =c (*OR* v advprt)) ;; cat value check
       (%(x0 obj) = *DEFINED*) ;; check if obj is defined
       (%x2 == %(x0 obj)) ;; np becomes obj
       (*TRY* ((%x2 case) = *UNDEFINED*) ;; if obj's case is undef
           ((%x2 fake-neg) =c +)) ;; if the higher verb is negated
           ((%x2 case) = g)) ;; assign gen
       ((%x0 neg) =c +) ;; if the verb is negated
       (%(x2 case) = g)) ;; assign gen
       ((%x2 case) = a)))) ;; else, assign acc
       (%x1 == %x0))) ;; v assigned other vp features
\end{verbatim}
Note that statements assigning genitive are ordered: negation of a higher verb (fake-neg) is checked first and if it fails, negation of the local verb (neg) is verified. This means that genitive is assigned only once: if the first check succeeds, the second one is not tried. This modification allows us to correctly resolve case in (14)-(15).

5 Aspect

Unlike in English, aspectual marking is incorporated into verbal morphology in Polish. In particular, this means that aspect is present on all verb forms, e.g., infinitives or gerunds, and not only on tensed forms. In our MT system, only perfective vs. imperfective distinction is considered and we deal with minimal aspectual pairs.\footnote{For an approach which incorporates more complex spatiotemporal relations (Aktionsarten) into an English-Polish MT system see Gawrońska (2001).}

In the system, lexical concepts corresponding to English verbs are ambiguous, i.e., they can be translated by either a perfective or an imperfective verb, see (10) and (11). The appropriate form is chosen according to several ordered context-sensitive heuristic rules specified in the mapper module. This means that the aspect value, introduced into FS, is resolved based on the interlingua (IR) specification provided by the system. Aspect assignment rules are discussed below.

5.1 Declarative Mood

For finite verbs in declarative mood, aspect assignment is primarily based on tense specification. In the system, all continuous forms, marked as (progressive +) in IR, are translated as imperfective. Next, forms of perfective tenses, i.e., (perfective +), are translated as perfective. Then, verbs in simple past or future simple tenses (marked as (tense past) or (tense future) in IR) are translated as perfective. This assignment is compatible with that proposed in Gawrońska (1993).

Additionally, we assume that certain types of subordinate conjunctions, e.g., ‘while’, ‘once’, ‘before’, etc., impose aspect requirements on a verb in the subordinate clause. For the current system, the following assignments have been proposed:

- ‘while’: imperfective

  (20) You can send an electronic fax \textbf{while} the printer makes copies.
  Można wysłać elektroniczny faks, podczas gdy drukarka robi k"opie.
  
  \textit{can} _{\textit{impers}} \textit{send} _{\textit{inf}} \textit{electronic} fax \textit{while} \textit{printer} \textit{makes} _{\textit{imperf}} copies

- ‘once’, ‘after’, ‘before’, ‘in order to’, ‘until’: perfective; additionally clauses introduced by conjunction ‘until’ have to be negated in Polish

  (21) Jobs also queue and wait \textbf{until} another job finishes.
Zadania także ustawiają się w kolejce i czekają, dopóki inne zadanie nie skończy się, not finishes REF

- 'by'+gerund: imperfective; such clauses are translated into Polish by a contemporary adverbial participle derived only from imperfective verbs, see Saloni and Świdziński (1985)

(22) Close the document by selecting <content> Close </content> from the <content> File </content> menu.
Zamknij dokument wybierając <content> Zamknij </content> z close document selecting<imperf> <content> Close </content> from menu <content> Plik </content>, menu <content> File </content>

If none of the above cases hold, we assume that aspect of present tense verbs is imperfective. This assignment is valid also for gerunds as they are represented in IR as present tense verbs with an additional feature (nominal +).
In order to assign aspect in sentences such as (23)-(24), more subtle rules have to be proposed.

(23) If the document is printed on a regular weight paper, ...

 passive: Jeśli dokument jest drukowany na papierze o standardowej gramaturze...
    if document is printed<imperf> on paper of standard weight...
       state: Jeśli dokument jest wydrukowany na papierze o standardowej gramaturze...
        if document is printed<perf> on paper of standard weight...

(24) If the document is printed by the printer, ...

    Jeśli dokument jest drukowany przez drukarkę ...
     if document is printed<imperf> by printer ...

In IR, both sentences are represented as passive constructions, distinguished only by the lack / presence of an overt argument. However, as noted in Crestan (2000), the difference is more subtle. (24) is a genuine passive and can be rendered by the active sentence: “The printer prints the document.” On the other hand, sentences such as (23) are semantically ambiguous: the participle can express a state or be a passive (agentless) form. Although in English this distinction is not overtly expressed, other languages, e.g., Polish, (23), or French, see Crestan (2000), use separate forms to capture the difference. In Polish, usually a different aspect is used, see (23), whereas in French a different tense is used. Although (23) might suggest that in Polish perfective results in a state reading, this is not always the case, see (25).

(25) The fax is stored in memory.

    state: Faks jest przechowywany w pamięci.
     fax is stored<perf> in memory
passive: Faks jest przechowywany w pamięci.

fax is stored_{perf} in memory

As discussed at length in Gawrońska (1993), aspect depends on semantic-pragmatic conditions (aspectual perspective). More precisely, it has to be determined whether the event / situation can be considered a distinct, integrated whole (perfective aspect) or not (imperfective).

Since the IR representation does not provide enough information to distinguish between state and passive readings, the correct distinction cannot be made for Polish (or French). In the Polish system, sentences as in (23) and (25) are assigned perfective aspect, which usually results in a state interpretation.

5.2 Imperative Mood

We assume that in imperative mood, aspect depends on negation. This observation is confirmed by a brief analysis of Polish technical documentation. We have examined 3 short instructions for changing ink cartridge, available from www.balta.pl/katalog/file (SC005FC3.jpg (Doc1), SH649FC3(51625).jpg (Doc2), SH625FC3.jpg (Doc3)). The documents contained, respectively, 32, 37 and 28 imperative verbs. The summary of obtained results is presented in fig. 2.

<table>
<thead>
<tr>
<th>ASPECT</th>
<th>NEGATED</th>
<th></th>
<th></th>
<th>NON-NEGATED</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Doc1</td>
<td>Doc2</td>
<td>Doc3</td>
<td>All</td>
<td>Doc1</td>
<td>Doc2</td>
</tr>
<tr>
<td>imperfective</td>
<td>6.25%</td>
<td>5.4%</td>
<td>3.57%</td>
<td>5.15%</td>
<td>9.38%</td>
<td>21.62%</td>
</tr>
<tr>
<td>perfective</td>
<td>3.12%</td>
<td>0%</td>
<td>0%</td>
<td>1.03%</td>
<td>81.25%</td>
<td>72.97%</td>
</tr>
</tbody>
</table>

Figure 2: Statistics for imperative forms in a sample of technical documentation

As these results indicate, in technical documentation, negated imperatives more often appear with imperfective forms (5.15%), whereas perfective aspect prevails with non-negated imperatives (78.35%).

Heuristic rules used in the system conform with the above statistics: we translate non-negated imperatives as perfective, (26a), and negated imperatives as imperfective verbs, (26b).

(26) a. Print a test page.
    Wydrukuj stronę próbną.
    print_{perf} page test

    b. Do not move the lever after the scanner has begun sending the page.
    Nie przesuwan dźwigni, gdy skaner zaczął wysyłanie strony.
    not move_{imperf} lever when scanner started sending page

This assignment, however, is not perfect as examples below indicate:
(27) Keep in mind that this technology is not perfect.
Pamiętaj, że ta technika nie jest doskonała.
remember<sup>imperf</sup> that this technology not is perfect

(28) (When you are using the OCR software, follow these guidelines in order to obtain the best possible results:)
Process only pages that have crisp, clear text.
Przetwarzaj tylko strony, które mają czytelny, wyraźny tekst.
process<sup>imperf</sup> only pages which have crisp clear text

In (27), imperfective is correctly obtained by the system because ‘keep in mind’ is considered a fixed phrase and, hence, exceptionally translated by an imperfective verb. In (28), however, no such assumption can be made. In this case, imperfective aspect is triggered by the context in the preceding sentence which introduces a sequence of general statements. Moreover, as (29) indicates, aspect assignment can be influenced by the semantic context introduced three sentences earlier:

(29) (When you are using the OCR software, follow these guidelines in order to obtain the best possible results:
Process only pages that have crisp, clear text.
Letters that have gaps, letters that bleed along their edges, or letters that touch other letters will not work well with the OCR software.
Note that underlined text and text that is close to non-text items, such as graphics, will also degrade OCR results.)
Process only pages that contain at least 9-point text.
Przetwarzaj tylko strony, które zawierają co najmniej dziewięciopunktowy tekst.
process<sup>imperf</sup> only pages that contain at least 9-point text

These examples indicate that in order to correctly resolve aspect in Polish translations, a deeper semantic representation is required (aspectual perspective, cf. Gawrońska (1993)). Such a detailed semantic representation is not available in KANTOO and, hence, the system cannot assign the correct imperfective form in (29) (or (28)); the perfective form przetwórz is used instead.

5.3 Infinitives

Infinitives have no mood or tense specified so we need separate rules to resolve aspect for these forms. In the English test corpus, infinitives appear as either complements of other verbs, e.g., modals, or they head infinitive clauses introduced by a conjunction such as ‘in order to’. We assume that in the former case, aspect of the infinitive depends on the governing verb while in the latter — on the subordinate conjunction.
Only some verbs such as ‘start’, ‘finish’ or ‘continue’ provide a clear aspect specification for their infinitival complement and require an imperfective verb in Polish. Other verbs can occur with both forms and the choice of aspect on their complements depends on the
situation / event semantics, see Gawrońska (1993). Since such a deep semantic representation is unavailable in the IR, we propose several tentative heuristic rules based on behaviour of infinitives in the English corpus.

For the conjunction 'in order to', we assume that it requires a perfective infinitive argument (see § 5.1). Aspect assignment for complements of modals is more complex. Modal verbs are represented in IR by a set of binary semantic features such as ability, possibility, tentativity, necessity, obligation, see Leavitt et al. (1994). The following aspect assignment has been adopted for complements of modals:

- 'can': (ability +) or (possibility +) → perfective

(30) You can also scan a document while another document is printing.
Można także zeskanować dokument, podczas gdy inny dokument drukuje can also scan_{perf} document while another document prints się.

REFL

- 'cannot': (ability +) (negation +) → imperfective

(31) You cannot scan documents.
Nie można skanować dokumentów.
not can scan_{imperf} documents

- 'cannot': (possibility +) (negation +) → perfective

(32) The printer cannot fax the page which is jammed.
Drukarka nie może przefaksować strony, która jest zakleszczona.
printer not can fax_{perf} page which is jammed

- 'could': (possibility +) (tentativity low) → perfective;

(33) The printer could not detect a dial tone.
Drukarka nie mogła rozpoznać sygnału.
printer not could detect_{perf} tone

- 'may': (possibility +) (tentativity medium) → perfective;

(34) It may not be connected properly to the computer.
Może nie być właściwie podłączony do komputera.
may not be properly connected_{perf} to computer

- 'must': (obligation medium) → perfective;

(35) You must unhook the other device in order to connect the printer.
Trzeba wyłączyć inne urządzenie, aby podłączyć drukarkę.
need unhook_{perf} another device in order to connect_{perf} printer

- 'should': (expectation +) → perfective;
The printer should work with this software.

Drukarka powinna działać z tym oprogramowaniem.
printer should work<sub>imperf</sub> with this software

(36)

Again, this method is not flawless as the choice of aspect strongly depends on semantic context. Let us consider the following example:

(37) (This product is specifically designed to allow you to do many tasks simultaneously.)

For example, you can print a document while you send a fax.

a. Na przykład, można drukować dokument podczas wysyłania faksu.
on example can print<sub>imperf</sub> document while sending fax

b. Na przykład, można wydrukować dokument podczas wysyłania faksu.
on example can print<sub>perf</sub> document while sending fax

If taken out of context, both (37a) and (37b) can be translations of the English sentence: the former is focused more on simultaneity of the two tasks whereas the latter indicates that one task is completed while the other is performed. If we take into account the previous context, however, the imperfective form in (37a) is preferred by native speakers. This contextual information is unavailable to the system and, hence, (37b) is generated instead.

6 **Personal vs. Impersonal Forms**

Polish is a pro-drop language which means that the nominative subject can be dropped but agreement features (number, person, gender) are still present on the verb. In fact, the pronominal subject is usually overtly unrealized unless it is required for the contrastive stress. In the system, the third person pronominal subject is systematically omitted in translation; in (38), the pronominal subject 'it' is not translated but the agreeing 3rd person singular verb form is still used.

(38) It may not be connected properly to the computer.

Może nie być właściwie podłączony do komputera.
can<sub>3_s</sub> not be<sub>m</sub> properly connected to computer

Moreover, in Polish technical documentation, addressing the reader directly should be avoided. Therefore, if possible, second person pronouns, i.e., 'you' or 'your', should be omitted in Polish translations and impersonal (non-agreeing) verb forms should be used instead.

In the English corpus, 'you' is usually used as the subject with modal verbs ('can', 'may', 'must', 'could') or in subordinate clauses introduced by 'while', 'if', 'once', 'after', 'until', 'when'. Below, we present several context-sensitive rules for modal verbs and subordinate conjunctions which have been proposed to eliminate direct translations of 'you'.

---

9A few exceptional rules have been additionally proposed to deal with second person forms of the verbs 'have', 'want', 'need' (see §8).
order to make the Polish output more natural, second person possessive pronouns are omitted in translation.

- 'can/may': ‘you can’ translated by the impersonal (non-agreeing) verb *można*, (39); otherwise: a personal (agreeing) form of the modal verb *móc*

  (39) You can send an electronic fax.
  Można wysłać elektroniczny faks.
  can_{impers} send_{mf} electronic fax

- 'must': ‘you must’ translated by the impersonal (non-agreeing) verb *trzeba*, (40); otherwise: a personal (agreeing) form of the modal verb *musić*

  (40) You must unhook the other device.
  Trzeba wyłączyć inne urządzenie.
  need_{impers} unhook_{mf} other device

- ‘could’: ‘you could’ translated by the impersonal (non-agreeing) conditional form *możnaby*, (41); otherwise: a conditional form of (agreeing) modal *móc*

  (41) You can send documents that you could not send otherwise:
  Można wysłać dokumenty, których nie można wyślij
  can_{impers} send_{mf} documents which not could send_{mf}

  w przeciwnym wypadku:
  otherwise:
  
- 'while S': ‘while you’+verb translated as *podczas+gerund*, (42); otherwise: *podczas gdy+S*;

  (42) You can print a document while you send a fax.
  Można wydrukować dokument podczas wysyłania faksu.
  can_{impers} print_{mf} document during sending fax

- 'if S': ‘if you’+verb translated as *w przypadku+gerund*, (43); otherwise: *jesli+S*;

  (43) If you are loading a business card, place it with a long edge first.
  W przypadku wprowadzania wizytówki umieść ją długą krawędzią do
  in case loading business card place it long edge to
  przodu.
  front

- ‘after/once S’: ‘after/once you’+verb translated as *po+gerund*, (44); otherwise: *gdy/gdy tylko+S*;

  (44) The printer finishes the rest of the job, once you have removed the jam.
  Drukarka kończy resztę zadania po usunięciu blokady.
  printer finishes rest job after removing jam
For other cases, no general substitution pattern could have been proposed. Therefore, the pronoun is dropped but an agreeing verb form still has to be used, which makes translations less natural.

(45) Make sure that you use a carrier sheet when you are copying fragile items and photographs.

Pamiętaj, by użyć opakowania ochronnego, kiedy kopiujesz delikatne dokumenty i zdjęcia.

and photographs

7 Middle Voice

Let us consider the following example:

(46) You can also scan a document while another document is printing.

Można także zeskanować dokument, podczas gdy inny dokument drukuje się.

REFL

The English form ‘is printing’ is translated by the so-called middle form drukuje się (lit.: ‘prints itself’). This form differs from the active voice as the object is not present and also from the passive voice as the active form drukuje is used. Note that drukuje is not a reflexive verb, either, i.e., it normally does not occur with the reflexive particle się.

Middle voice is not very productive in Polish. Hence, all verbs which allow for this form need a double lexical specification as either regular transitive verbs (i.e., they require an NP object) or as middle forms (i.e., they require a reflexive particle).

However, there is yet an additional complication brought by transitive verbs with an elided object, (47).

(47) Press the key when you are printing.

a. Naciśnij przycisk kiedy drukujesz.
   press key when print_{2,sg}
   b. *Naciśnij przycisk kiedy drukujesz się.
   press key when print_{2,sg} REFL

In fact, in IR the verb ‘printing’ in (46) and (47) has the same semantic representation and the two sentences are indistinguishable. However, as Polish translations indicate, only the former is a middle form and should occur with the reflexive particle się. In order to avoid illicit (47b), semantic constraints on subjects of middle and active voice forms of specific verbs have to be added. For middle voice of ‘print’, the subject can be a printable entity, e.g., a file, a document, a fax, but not a personal pronoun; for elided constructions, however, the subject can be a printing entity, e.g., a printer, a fax machine, a person, but
not a document, a fax, etc. At present, this distinction has been implemented by explicitly enumerating specific printable/printing entities. For a fully developed system, however, a general semantic specification is required.

8 Word Order

One of the problems which has to be addressed while dealing with Polish generation is word order. In Polish, word order is not constrained by grammatical functions but rather by discourse factors and information structure. Since representation of this information is quite restricted in the KANTOO system, our account of word order in the Polish module will be rather limited. A few issues are addressed below.

8.1 Reversed Subject

The neutral word order in Polish is subject-verb-object (SVO) but all six permutations are possible. As noted in Siewierska (1988) and Styś (1998), the two most common and least emphatic word orders in Polish are SVO and VOS. In the current system prototype, the former is adopted as a basic order, whereas the latter is used only for a few special constructions, e.g., translations of ‘you may need’+VP[inf] or existential sentences with the non-referential there.

‘You may need’ As discussed in §6, second person modals are normally translated by impersonal verbs. The ‘you may need’+VP[inf] phrase, however, is best translated as może wystąpić potrzeba+gerund (lit.: ‘may3.sg occur a need (for)’+gerund) where the verb może is a genuine modal agreeing with the reversed nominal subject potrzeba ‘a need’. In this expression, the subject has to be postverbal.

The entire expression is stored as a phrase in the lexicon, with the subject marked as (subj-final +), indicating the postverbal position. (48).

(48) wyst/api potrzeba = ((lex wyst/api/c-v)
    (subj ((lex potrzeba)
        (subj-final +)
        (agr ((number singular)))))),

A separate syntactic rule is added to the grammar to account for sentences with the reversed subject. An example of its performance is given in (49).

(49) You may need to add another parallel interface card.

Może wystąpić potrzeba dodania równoległąj karty sprzegającej.

Another type of constructions which require the reversed subject in a Polish translation are existential sentences with there. Following Styś (1998), such constructions are translated adopting the pattern:
there+be+NP(+location-PP) → (location-PP+)być_verb+NP

where the final NP in translation is the subject agreeing with the preceding verb. More precisely, a mapper rule ensures that semantic arguments of the predicate *A-BE-EXISTENCE (which represents the existential verb ‘be’ in the IR) are assigned the appropriate syntactic functions in Polish. Additionally, since the subject is required to be postverbal, it is marked (subj-final +). The grammar rule which allows for postverbal subject realisation guarantees the correct surface order, (50).

(50) There may be a poor phone connection.
    Może być zło połączenie telefoniczne.
    can₃sg beₘₙf bad connection phone

If these two types of constructions are combined, they give the following result:

(51) You may also need to clean these parts if there are smudges or other marks on
documents or on scans of documents.
    Może także wystąpić potrzeba oczyszczenia tych części, jeśli na dokumentach lub
    can₃sg also occurₘₙf need cleaning these parts if on documents or
    na skanach dokumentów są smugi lub inne zabrudzenia.
    on scans documents are smudges or other dirt

Passive Sentences English passives are often translated into Polish as an active sentence with the reversed order of grammatical functions, i.e., OVS order results, (52), or are rendered by impersonal forms if no by-phrase is present, (53), see Styś (1998) for details.

(52) The fax has been received by the printer. →
    Faks odebrała drukarka.
    fax received printer

(53) OCR software, which **was installed** with your other software →
    oprogramowanie OCR, które **zainstalowano** razem z innym
    software OCR which **installed**ören_prt together with other
    oprogramowaniem

There are two reasons, however, for which we decided to retain passive in Polish translations. First, inversions such as in (52), seem to be less acceptable in certain cases and a passive form is preferred:

(54) Text cannot be edited by a **word-processor program.** →

a. *Tekstu nie można wyedytować **program przetwarzania tekstów.***
    text not can edit program processing texts

b. *Tekst nie może być wyedytowany przez **program przetwarzania tekstów.***
    text not can beₘₙf edited by program processing texts
The problem might be due to the 'heaviness' mismatch between the object 'text' and the subject 'word-processor program' or inanimacy. The second, and more important, reason is related to translations of English agentless passives by impersonal forms, as in (53). These impersonal forms have (inherently) past meaning and, therefore, cannot be used for translation in other tenses, (55).

(55) A fax is being received. →

a. Faks jest odbierany.
   fax is received
b. #Odbierano faks.
   receivedimpers_prt fax

For these reasons, we prefer to retain passive in Polish translations.

8.2 Sentential vs. Verbal Adjuncts

In the system, two kinds of verbal adjuncts are distinguished: verbal (attached directly to the verb) and sentential (at the sentence level). There are several features in IR which indicate various types of modification, e.g., manner for adverbs, q-modifier for PP modifiers or qualification for subordinate clause modifiers. In IR, these modifiers are specified as attributes of the main verbal predicate, (56), although their syntactic attachment can be different.

(56) OCR works correctly if the text is clear.

(*A-WORK
 (agent
   (*O-OPTICAL-CHARACTER-RECOGNITION
    (acronym +)
    (implied-reference +)
    (number mass)))
 (argument-class agent)
 (manner
   (*M-CORRECTLY
    (degree positive)))
 (mood declarative)
 (punctuation period)
 (qualification
   (*G-QUALIFYING-EVENT
    (event
     (*A-BE-PREDICATE
      (attribute
       (*P-CLEAR
       (agent
        (*O-OPTICAL-CHARACTER-RECOGNITION
         (acronym +)
         (implied-reference +)
         (number mass)))
 (argument-class agent)
 (manner
   (*M-CORRECTLY
    (degree positive)))
 (mood declarative)
 (punctuation period)
 (qualification
   (*G-QUALIFYING-EVENT
    (event
     (*A-BE-PREDICATE
      (attribute
       (*P-CLEAR
       (agent
        (*O-OPTICAL-CHARACTER-RECOGNITION
         (acronym +)
         (implied-reference +)
         (number mass)))
 (argument-class agent)
 (manner
   (*M-CORRECTLY
    (degree positive)))
 (mood declarative)
 (punctuation period)
 (qualification
   (*G-QUALIFYING-EVENT
    (event
     (*A-BE-PREDICATE
      (...)
(degree positive)))
(mood declarative)
(predicate-role attribute)
(tense present)
(theme
 (*C-TEXT
   (number
    (:CR mass singular))
   (reference definite)))
(extent
 (*CONJ-if)))
(tense present))

The classification into verbal vs. sentential adjuncts helps us to establish the order of adjuncts in the Polish output.

8.2.1 Verbal Adjuncts

Although the position of most adjuncts in Polish is relatively free, for some adverbs verb adjacent positions are preferred (or required, cf. Śpiwak and Szymańska (1995)).

(57) The characters have been correctly interpreted by the OCR software.
Znaki zostały poprawnie zinterpretowane przez oprogramowanie OCR.
characters were correctly interpreted by software OCR

To the best of our knowledge, an exhaustive list of such adverbials is not available. Hence, we decide arbitrarily, based on acceptance of an adverbial in a specific sentence position, whether to treat it as a verbal or a sentential modifier. If the adverbial can be easily placed in different positions without affecting the meaning, it is considered a sentential modifier, see §8.2.2. On the other hand, if distinct non verb adjacent positions either make the sentence ungrammatical or affect the meaning, we assume that the adjunct is a verbal modifier. In the system, such adjuncts have the attribute verbal specified in the lexicon. The value of this attribute, pre or post, indicates preferred position with respect to the verb. In (57), the position of popraniemie is obtained as desired since the adverb is marked (verbal post) (with respect to the main verb).
Certain verbal adverbs do not have a fixed position with respect to the verb.

(58) (Letters that have gaps, letters that bleed along their edges, or letters that touch other letters will not work well with the OCR software.)
Note that underlined text and text that is close to non-text items, such as graphics, will also degrade OCR results.
Zauważ, że podkreślenie i tekst, który jest zbliżony do nietekstowych obiektów note that underlying and text which is close to non-text items takich jak grafika także pogorszą jakość przetwarzania OCR, such as graphics also degrade quality processing OCR
(59) (This user guide provides details about performing tasks from the product control panel, changing infrequently used settings, resolving problems, and contacting <trademark>HP</trademark> support.)

This guide also provides an overview of software functions.

Ten podręcznik zawiera także przegląd funkcji oprogramowania. This guide contains also overview functions software

In (58), także ‘also’ occurs preverbally, whereas in (59) — postverbally. Note that position of the adverb is not optional, as changing its position affects sentence meaning. If także follows the verb in (58), the sentence meaning shifts to ‘also OCR results will be degraded’ (and possibly something else), which is incorrect according to the context in the previous sentence. On the other hand, preverbal także in (59) makes the sentence ambiguous as the meaning ‘also this guide (just like some other guides) provides information’ is added, which is not the sense conveyed by the English sentence. Hence, the adverb position is determined by semantics rather than by purely syntactic factors. Although in the above two examples, preverbal position for także could be adopted in the system (this position ensures the correct interpretation to be obtained in both cases), this solution is not flawless. Position of także seems to additionally depend on information structure as the adverb is an emphasis / focus marker. This is reflected by the fact that its topicalized positions are normally disallowed (focus elements cannot occur in initial position in Polish).

(60) Also, try connecting a phone directly to the phone line.

a. Spróbuj także podłączyć telefon bezpośrednio do linii telefonicznej.

b. *Także próbow podłączyć telefon bezpośrednio do linii telefonicznej.

Since the necessary pragmatic information is unavailable in the IR, we have to mimic conditions on the adverb placement in syntax. In the lexicon, także is specified as a verbal adverb and is assigned postverbal position, i.e., (verbal pcst). This allows us to rule out (60b) and correctly render (60a) and (59), as desired. However, the most preferred preverbal position in (58) is not obtained.

8.2.2 Sentential Adjuncts

Since all sentential adjuncts have the same syntactic function and are attached at the same syntactic level, we represent all types of (non-verbal) adjuncts by a single v-mod attribute; hence, its value is a list. The order of elements on this list is determined by 1) the order in which IR attributes are mapped into f-structure (the following order has been adopted: manner, q-modifier, qualification, adjoined-modifier) and 2) if there are several adjuncts of the same type, e.g., q-modifier, by their surface order. On this list, preposed adjuncts are also distinguished from postposed ones. Two recursive syntactic rules add sentential adjuncts: preposed adjuncts precede a sentence, whereas other adjuncts follow the sentence. (These are the two most frequent adjunct positions in Polish, see Siewierska (1993)).
(61) For example, if a fax is being received while you are printing a job, the fax is stored in memory.

Na przykład jeśli faks jest odbierany podczas drukowania zadania, faks jest na przykład, jeśli faks jest odbierany podczas drukowania zadania, faks jest on example if fax is received during printing job fax is stored in memory.

In (61), there are three adjuncts: ‘for example’, ‘if’-sentence and ‘in memory’. These adjuncts are represented in IR as values of adjoined-modifier, qualification and q-modifier, respectively. The reversed mapping order of the two preposed adjuncts matches their surface order (‘for example’ precedes the ‘if’-sentence) and the order of adjuncts in the output is the same as in the input. There is only one postverbal adjunct (‘in memory’) and it is put at the end of the sentence.

A problem appears if there is a mismatch between surface and (reversed) mapping ordering, see (62a) and (62c), which is usually due to wrong classification of an adjunct.

(62) a. Alternatively, if the test page was too light, check whether the toner cartridge needs to be replaced.

b. Ewentualnie, jeśli strona próbna była zbyt jasna, sprawdź, czy alternatively if page test was too light check whether pojemnik z tonerem nie powinien zostać wymieniony.

c. *Jeśli strona próbna była zbyt jasna, ewentualnie sprawdź, czy pojemnik z tonerem nie powinien zostać wymieniony.

In IR, the adjunct ‘alternatively’ is (wrongly) identified as manner, whereas the ‘if’-sentence is represented as qualification, and they are mapped onto v-med in this order. This forces their reversed surface realisation, which is incorrect, see (62b) vs. (62c).

The easiest way to deal with such mismatches is to retain multiple representation of adjuncts in f-structure and apply separate ordering rules to distinct adjunct types.

9 Pronominal Anaphora Resolution

As mentioned in §2, in the current system 4 genders for Polish are considered: masculine, masculine-human, neuter and feminine. Therefore, in translating English pronominal anaphora to Polish, an appropriate gender distinction has to be made. The KANTOO analyser allows for pronominal anaphora resolution and information about the antecedent is present in the IR, (63).
(63) Use the latches on the printer door to pull it open.

(*A-USE
 (argument-class agent+patient+complement)
 (complement
  (*G-COMPLEMENT-EVENT
   (event
    (*A-PULL
     (agent
      (*O-LATCH
       (gapped +)
       (number plural)
       (q-modifier
        (*Q-located_ON
         (case
          (*K-ON))
         (object
          (*O-PRINTER-DOOR
           (number singular)
           (reference definite)))
         (role located)))
       (reference definite)))
      (argument-class agent+theme)
     (manner
      (*M-OPEN
       (degree positive))))
     (theme
      (*G-PRONOUN
       (antecedent
        (*C-PRINTER-DOOR)
        (number singular)
        (person third)
        (reference definite))))
     (extent
      (*CCNJ-infinitive))))
    (mood imperative)
   (patient
    (*O-LATCH
     (number plural)
     (q-modifier
      (*Q-located_ON
       (case
        (*K-ON))
       (object
       (reference definite)))))
    (extent
     (*CCNJ-infinitive))))
  (role located)))
 (reference definite)))
 (argument-class agent+theme)
 (manner
  (*M-OPEN
   (degree positive)))
 (theme
  (*G-PRONOUN
   (antecedent
    (*C-PRINTER-DOOR)
    (number singular)
    (person third)
    (reference definite)))
  (extent
   (*CCNJ-infinitive))))
(mood imperative)
(patient
 (*O-LATCH
  (number plural)
  (q-modifier
   (*Q-located_ON
    (case
     (*K-ON))
    (object
    (reference definite)))))
Użyj zacisków na drzwiczkach drukarki, by otworzyć je.

The algorithm of pronominal anaphora resolution takes into account the current and the preceding sentences only, which in a general case is insufficient (also for Polish, see Mitkov and Styś (1997)). However, for technical texts written in controlled language, this does not seem to be a limiting factor, see Mitamura et al. (2002). Preference heuristics used in the KANTOO anaphora resolution algorithm are specified below (for details see Mitamura et al. (2001), Mitamura et al. (2002)):

- prefer an antecedent that is also an anaphor
- prefer an antecedent that is not a (SGML) tag
- if two possible antecedents form a phrase "NP1 of NP2", prefer NP1, unless NP1 is "time", "type", "length", "size", "part" or "quality"
- prefer antecedents that attach to the same syntactic constituent as the pronoun
- prefer antecedents that attach to the same part of speech as the pronoun
- prefer antecedents that have the same syntactic function as the pronoun
- prefer antecedents that are conjuncts
- prefer nominal antecedents that have a determiner, quantifier or possessor, or are the value of a (SGML) tag
- prefer nominal antecedents that have a definite determiner
- prefer the most recent antecedent

Since the antecedent is already indicated in IR, a mapping rule has to ensure only that the pronoun gets the gender of its antecedent (for a single concept, e.g., *C-PRINTER-DOOR in (63), the gender is encoded directly in its lexical entry; for conjoined antecedents, the gender depends on the conjuncts, see §9.2).

The results of anaphora resolution algorithm obtained for Polish in the tested text sample are discussed below.
9.1 Subject Pronouns

Polish is a pro-drop language and, as mentioned in §6, subject pronouns are dropped in translation. Nevertheless, due to subject-verb agreement, morphological features of the nominative subject (gender, person, number) are present on the verb. Additionally, predicative adjectives or participles in passive agree in gender (and number) with the subject, (64).

(64) A fax is being received while you are printing a job.
Faks jest odbierany podczas drukowania zadania.
fax_{msc} is received_{msc} during printing job

These two agreement types have been used to verify gender selection when the subject pronoun is dropped.

In the test corpus, there were 9 instances of ‘it’ used as a subject pronoun (3 of them occur in sentences the Polish system cannot handle yet) and 2 of ‘they’-subject. Among these sentences, 4 shown subject-participle agreement and in all 4 cases the form of the participle indicated that the antecedent’s gender has been correctly specified.

As for the subject-verb agreement in Polish, overt gender distinction appears only on verbs in past tense, see §2 (and in some future tense forms). Since none of the subject-dropped sentences had a verb in past form, gender selection was not confirmed in these cases.

The summary of results for subject pronouns is presented in (65).

<table>
<thead>
<tr>
<th>(65)</th>
<th>it-subject</th>
<th>they-subject</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>translated (correct)</td>
<td>6 54.6%</td>
<td>2 18.2%</td>
<td>72.8%</td>
</tr>
<tr>
<td>a) gender on participle</td>
<td>2 18.2%</td>
<td>2 18.2%</td>
<td>36.4%</td>
</tr>
<tr>
<td>b) no gender shown</td>
<td>4 36.4%</td>
<td>0 0%</td>
<td>36.4%</td>
</tr>
<tr>
<td>no translation</td>
<td>3 27.2%</td>
<td>0 0%</td>
<td>27.2%</td>
</tr>
</tbody>
</table>

9.2 Object Pronouns

For object pronouns, not only the correct gender distinction has to be made but they have to appear in the appropriate case specified by the verb.

Among the tested examples, only one incorrect realisation has been found, see (66) (the antecedent chosen by KANTOO is indicated in boldface, whereas the correct antecedent is underlined).

(66) If you use this procedure, then the setting remains on for all jobs until you change it again.
W przypadku stosowania tej metody ustawienie pozostaje włączone in case applying this method fem\_{gen} setting_{nt,gen} remains switched on dla wszystkich zadań, dopóki ponownie nie zmienisz jej.
for all jobs_{nt,gen} untill again not change it_{fem,gen}
The wrong translation is due to incorrect anaphora resolution in the IR: ‘procedure’, rather than the correct ‘setting’, is specified as the antecedent for the pronoun ‘it’ (see Mitamura et al. (2002) for discussion). Since ‘procedure’ and ‘setting’ have distinct genders in Polish (the former is feminine, whereas the latter is neuter) and the corresponding pronouns have different case forms, *jej* vs. *go*, the translation is incorrect.

For other occurrences of object pronouns, gender was successfully specified, also for cases of interclausal anaphora, cf. (67).

(67) (Near the top, inside edge of the document release door, inspect the three pressure feet and the pad for dirt or debris.)

If they are dirty, clean them with a lint-free cloth sprayed with a small amount of alcohol.

Jeśli są brudne, oczysć je tkaniną bezwłóknową, która jest nawiłżona niewielką ilością alkoholu.

In (67), the antecedent of ‘they’ and ‘them’ is the coordinated phrase ‘the three pressure feet and the pad’ in the preceding sentence. This phrase is correctly picked by the anaphora resolution algorithm. As the antecedent is a coordinated phrase, some explanation of gender and number assignment in coordination in Polish is necessary. First, we assume that Polish coordinated NPs are plural. (This is not always the case, e.g., Kallas (1974)). Second, since NP conjuncts usually have distinct genders, gender of the entire coordinated phrase has to be specified separately. As shown in §2, Polish personal verbs have different gender forms in plural and singular. In short, three gender classes can be specified in singular (feminine, neuter, while masculine and masculine-human are grouped together), whereas there are only two in plural (masculine-human, ‘mhum’ vs. all remaining genders, which we call for convenience non-masculine-human, ‘non-mhum’).

We adopt here a somewhat simplistic approach of Szpakowicz (1986) and assign ‘mhum’ to a coordinated NP if one of its conjuncts is ‘mhum’; otherwise, ‘non-mhum’ is assigned. In (67), none of the conjuncts in the antecedent is ‘mhum’ (*pedal ‘pressure foot’ is masculine, whereas podkładka ‘pad’ is feminine). Therefore, the coordinated NP has gender ‘non-mhum’, which is assigned by the mapping rule to the pronouns (‘they’ and ‘them’).

In the test corpus, ‘it’ occurred 10 times as an object, whereas ‘them’ appeared as an object 3 times. The reference of ‘it’ was incorrectly resolved once, (66). while all instances of ‘them’-object were correct.

The summary results for object pronouns are given in (68).

(68) |            | it-object | them-object | All |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
</tr>
<tr>
<td>correct</td>
<td>9</td>
<td>69.2%</td>
<td>3</td>
</tr>
<tr>
<td>incorrect</td>
<td>1</td>
<td>7.7%</td>
<td>0</td>
</tr>
</tbody>
</table>
10 Comparison

In this section we present a brief comparison of our system with two English-Polish ma-
chine translation systems available on-line: Poltran (www.pcltran.com) and Matchpad
(www.systransoft.com/R&D/Matchpad/). Since the scope of translation in the three sys-
tems is different, their lexical coverage varies; KANTOO and Matchpad are domain-specific
(printer manuals vs. administrative texts), whereas Poltran is a general purpose system.
Moreover, their overall goals differ as well: the English-Polish module in KANTOO is meant
as a research prototype (it has a small lexicon of about 480 lexical entries and a generation
grammar consisting of 97 phrase rules and additionally about 1000 rules for processing SGML
tags), whereas Matchpad (still under development) and Poltran are supposed to be publicly
available fully developed systems. Due to these differences, our comparison will focus on
linguistic coverage rather than on accuracy of translations.

For evaluation purposes, a small test suite of 50 phrases has been created, see Appendix
C. The test suite covers the following phenomena: case assignment (by verbs, prepositions,
nouns and in relative clauses), aspect, word order (reversed subject and adjunct placement)
and anaphora resolution (subject and object pronouns, and interclausal anaphora). Some
aspects of translation discussed earlier on in the paper are not included in the test suite
as they are specific for technical documentation (personal vs. impersonal form) or they are
rather restricted (middle voice).

Since performance of KANTOO has been discussed in detail in the previous sections, we
concentrate here on analysis of results obtained on the test suite for Poltran and Matchpad.

10.1 Case Assignment

Case assignment in Poltran and Matchpad is rather limited. None of them implement genitive
of negation (GoN): neither in simple sentences, VC or relative clauses (the relative pronoun
should appear in genitive if the verb in the relative clause is negated). In both systems,
accusative seems to be a default case assigned by verbs. As for case assigned by prepositions,
Matchpad misses some prepositions or skips them in translation but the case assigned to NP
is usually correct. Complements of nouns (gerunds) had case resolved correctly in Poltran.
Matchpad, on the other hand, does not support gerunds so this assignment could not have
been tested.

10.2 Aspect

Treatment of aspect in Poltran and Matchpad seem quite neglected. Matchpad uses the
same aspect for all forms generated from a verb, although different tense and mood forms
are provided. Poltran uses perfective only for perfective tenses, whereas imperfective is used
otherwise. For the future tense forms, Poltran and Matchpad use a different form, analytical
future tense (formed with an auxiliary być ‘be’ and an imperfective verb), than KANTOO
but all three systems provide correct forms. As for aspect in imperative mood, Poltran
and Matchpad always use imperfective. For single sentences, if no context is given, both
imperfective and perfective forms are correct with non-negated verbs, although in technical
documentation perfective appears more often; hence its use in KANTOO. Since Poltran and Matchpad assign imperfective to non-negated imperative verbs, translations of idiomatic ‘keep in mind’ and context-dependent aspect in #24 are unproblematic for these systems, unlike for KANTOO. Poltran and Matchpad do not provide any special rules for aspect in subordinate clauses and general rules are used.

10.3 Word Order

Specific constructions with the reversed subject discussed in §8 do not deserve any special treatment in Poltran. In Matchpad, the correct verb form može is used but no additional order rules for the subject are applied. As for adjuncts placement, both systems rely on the word order in original text and translations retain the same adjunct positions as in English. This means that none of the systems (including KANTOO) get the postverbal position of ‘also’ in #41 but Poltran and Matchpad perform better with respect to sentential adjuncts placement as they retain English surface order.

10.4 Anaphora Resolution

Neither Poltran nor Matchpad provide any special treatment for pronominal anaphora. Poltran uniformly translates (both subject and object) pronoun ‘it’ by the neuter demonstrative pronoun to ‘this’. For plural pronouns, the system uses the pronoun one ‘they$_{non\_rhum}$’ but sometimes the masculine-human pronoun omi ‘they$_{rhum}$’ is provided as an option. On the other hand, Matchpad translates both singular ‘it’ and plural ‘them’ object pronouns by the singular masculine pronoun on. The subject pronoun ‘it’ is rendered by the singular demonstrative pronoun to ‘this’, whereas the plural subject pronoun ‘they’ is dropped in translation. None of the systems assigns case value to the pronoun and it always appears in nominative.
10.5 Summary Results

<table>
<thead>
<tr>
<th>sent. #</th>
<th>KANTOO</th>
<th>Poltran</th>
<th>Matchpad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>+,+</td>
<td>+,+</td>
<td>-,+</td>
</tr>
<tr>
<td>13</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>15</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>16</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>17</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>23</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>25</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>26</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>27</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>29</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>30</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>37</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
### (70)

<table>
<thead>
<tr>
<th>sent. #</th>
<th>KANTOO</th>
<th>Poltran</th>
<th>Matchpad</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>32</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>33</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>34</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>35</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>36</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>37</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>38</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>39</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>40</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>41</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>42</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>43</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>44</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>45</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>46</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>47</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>48</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>49</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>50</td>
<td>+</td>
<td>+,−</td>
<td>+,−</td>
</tr>
</tbody>
</table>

### 11 Conclusions

In this paper we presented a prototype of an English-Polish MT system. The system is designed to translate domain-specific technical texts written in controlled language. KANTOO uses a relatively rich interlingua representation, which allows us to deal with quite complex linguistic issues such as case assignment, aspect, word order or anaphora resolution. Although the generation rules developed in the system have been used for a specific application (translation of technical documentation), some of them are quite general and could be applied in other systems. For example, case assignment has been based on the theoretical study of Przepiórkowski (1999), whereas word order or aspect assignment have been (at least partly) derived from statistical analysis of Polish texts. Similar techniques can be applied to improve rules proposed in the system (e.g., aspect assignment on infinitives) and to provide further generalisations. This is left as a topic for further study.
A  Verbal Inflection

Classification of Polish verbs based on verb endings in third person singular present tense form (groups implemented in the system are indicated in the typewriter font):

<table>
<thead>
<tr>
<th>class</th>
<th>3rd sg present</th>
<th>group</th>
<th>3rd sg present</th>
<th>infinitive</th>
<th>1st sg past</th>
<th>present</th>
<th>imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-a</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2a</td>
<td>-uje</td>
<td>-ować</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2b</td>
<td>-uje</td>
<td>-ywać</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2c</td>
<td>-aje</td>
<td>-ować</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>-c</td>
<td>3a</td>
<td>-eje</td>
<td>-ać</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3b</td>
<td>-eje</td>
<td>-eć</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3c</td>
<td>-je</td>
<td>-ę</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4a</td>
<td>-nie</td>
<td>-nać</td>
<td>-nał</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4b</td>
<td>-nie</td>
<td>-nać</td>
<td>-nał</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4c</td>
<td>-nie</td>
<td>-nać</td>
<td>-l</td>
<td></td>
<td>-ńy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5a</td>
<td>-nie</td>
<td>-qć</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5b</td>
<td>-niej</td>
<td>-qę</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6a</td>
<td>-nić</td>
<td>-ać</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6b</td>
<td>-nić</td>
<td>-eć</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6c</td>
<td>-nić</td>
<td>-eć</td>
<td>-art</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7a</td>
<td>-ści, -śćć</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7b</td>
<td>-ć</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>-y(i)</td>
<td>9a</td>
<td>-eć</td>
<td>no stem alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9b</td>
<td>-eć</td>
<td>stem alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10a</td>
<td>-ć</td>
<td>no stem alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10b</td>
<td>-ć</td>
<td>stem alt.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### B Noun Inflection

#### B.1 Masculine Nouns

Classification of Polish masculine nouns based on the singular genitive form, stem alternations and vowel shifts.

<table>
<thead>
<tr>
<th>MASCULINE, NON-HUMAN</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>group in KANTOO</td>
<td>group in McShane</td>
<td>example</td>
<td>English</td>
<td>genitive singular</td>
<td>stem-final consonant</td>
</tr>
<tr>
<td>m1</td>
<td>1</td>
<td>papier</td>
<td>paper</td>
<td>-u</td>
<td>alternating</td>
</tr>
<tr>
<td>3</td>
<td>skaner</td>
<td>scanner</td>
<td>-a</td>
<td>alternating</td>
<td>no</td>
</tr>
<tr>
<td>m2</td>
<td>2</td>
<td>przewód</td>
<td>cord</td>
<td>-u</td>
<td>alternating</td>
</tr>
<tr>
<td>m5</td>
<td>5</td>
<td>plik</td>
<td>file</td>
<td>-u</td>
<td>non-alternating</td>
</tr>
<tr>
<td>7</td>
<td>pojemnik</td>
<td>container</td>
<td>-a</td>
<td>non-alternating</td>
<td>no</td>
</tr>
<tr>
<td>m6</td>
<td>6</td>
<td>dodatek</td>
<td>appendix</td>
<td>-u</td>
<td>non-alternating</td>
</tr>
<tr>
<td>8</td>
<td>nagłówek</td>
<td>heading</td>
<td>-a</td>
<td>non-alternating</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>ząb</td>
<td>tooth</td>
<td>-a</td>
<td>alternating</td>
<td>yes</td>
</tr>
<tr>
<td>9</td>
<td>krab</td>
<td>crab</td>
<td>-a</td>
<td>alternating</td>
<td>no</td>
</tr>
<tr>
<td>10</td>
<td>wół</td>
<td>ox</td>
<td>-a</td>
<td>alternating</td>
<td>yes</td>
</tr>
<tr>
<td>11</td>
<td>koń</td>
<td>horse</td>
<td>-a</td>
<td>non-alternating</td>
<td>no</td>
</tr>
<tr>
<td>12</td>
<td>jednorozec</td>
<td>unicorn</td>
<td>-a</td>
<td>non-alternating</td>
<td>yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MASCULINE-HUMAN</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>group in KANTOO</td>
<td>group in McShane</td>
<td>example</td>
<td>English</td>
<td>genitive singular</td>
<td>stem-final consonant</td>
</tr>
<tr>
<td>m17</td>
<td>17</td>
<td>odbiorca</td>
<td>recipient</td>
<td>-y</td>
<td>alternating</td>
</tr>
<tr>
<td>13</td>
<td>pasierb</td>
<td>stepson</td>
<td>-a</td>
<td>alternating</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>posel</td>
<td>deputy</td>
<td>-a</td>
<td>alternating</td>
<td>no</td>
</tr>
<tr>
<td>15</td>
<td>słuchacz</td>
<td>listener</td>
<td>-a</td>
<td>non-alternating</td>
<td>no</td>
</tr>
<tr>
<td>16</td>
<td>cudzoziemiec</td>
<td>foreigner</td>
<td>-a</td>
<td>non-alternating</td>
<td>yes</td>
</tr>
<tr>
<td>18</td>
<td>Krupski</td>
<td>(name)</td>
<td>adj.-like</td>
<td></td>
<td>no</td>
</tr>
<tr>
<td>19</td>
<td>Jerzy</td>
<td>(name)</td>
<td>adj.-like</td>
<td></td>
<td>no</td>
</tr>
<tr>
<td>20</td>
<td>Linde</td>
<td>(name)</td>
<td>adj-noun</td>
<td></td>
<td>yes</td>
</tr>
</tbody>
</table>

37
# B.2 Feminine Nouns

Classification of Polish feminine nouns based on the singular nominative form, stem alternations and vowel shifts.

<table>
<thead>
<tr>
<th>group in KANTOO</th>
<th>group in McShane</th>
<th>example</th>
<th>English</th>
<th>nominative sing. ending</th>
<th>stem</th>
<th>vowel shift</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1</td>
<td>1</td>
<td>grafika</td>
<td>graphics</td>
<td>-a</td>
<td>alt.</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>f3</td>
<td>3</td>
<td>tabela</td>
<td>table</td>
<td>-a</td>
<td>non-alt.</td>
<td>some</td>
<td></td>
</tr>
<tr>
<td>f5</td>
<td>5</td>
<td>opcja</td>
<td>option</td>
<td>{s,z,c}+ja</td>
<td>no</td>
<td>j does not merge with i</td>
<td></td>
</tr>
<tr>
<td>f7</td>
<td>7</td>
<td>loteria</td>
<td>lottery</td>
<td>{t,d,r,l,k, g,h,f}+ia</td>
<td>no</td>
<td>two i</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>kopia</td>
<td>copy</td>
<td>{b,p,m,n}+ia</td>
<td>no</td>
<td>two i</td>
<td></td>
</tr>
<tr>
<td>f12</td>
<td>12</td>
<td>odpowiedź</td>
<td>answer</td>
<td>consonant</td>
<td>no</td>
<td>-i endings</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>stopa</td>
<td>foot</td>
<td>a</td>
<td>alt.</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>aleja</td>
<td>alley</td>
<td>vowel+ja</td>
<td>no</td>
<td>j merges with i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ciocia</td>
<td>aunt</td>
<td>{s,z,c}+ia</td>
<td>no</td>
<td>one i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>cukierka</td>
<td>pastry shop</td>
<td>{b,p,m,n}+ia</td>
<td>no</td>
<td>native; one i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>bogini</td>
<td>goddess</td>
<td>ym/ini</td>
<td>no</td>
<td>derived from masc. nouns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>noc</td>
<td>night</td>
<td>consonant</td>
<td>some</td>
<td>-y endings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>kąpiel</td>
<td>bath</td>
<td>l</td>
<td>no</td>
<td>-i endings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>łódź</td>
<td>boat</td>
<td>consonant</td>
<td>yes</td>
<td>-i endings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>baronowa</td>
<td>baroness</td>
<td>-a</td>
<td>no</td>
<td>adj. paradigm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# B.3 Neuter Nouns

Classification of Polish neuter nouns based on the singular nominative form, stem alternations and vowel shifts.

<table>
<thead>
<tr>
<th>group in KANTOO</th>
<th>group in McShane</th>
<th>example</th>
<th>English</th>
<th>nominative singular ending</th>
<th>vowel shift</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>1</td>
<td>pismo</td>
<td>writing</td>
<td>alternating cons. +o</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>n2</td>
<td>2</td>
<td>szkło</td>
<td>glass</td>
<td>alternating cons. +o</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>n4</td>
<td>4</td>
<td>narzędzie</td>
<td>tool</td>
<td>any cons. +ie</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>n0</td>
<td>–</td>
<td>menu</td>
<td>menu</td>
<td>foreign; vowel</td>
<td>no</td>
<td>uninflected</td>
</tr>
<tr>
<td>3</td>
<td>oblicze</td>
<td>face</td>
<td>non-alternating cons. +{e,o}</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>muzeum</td>
<td>museum</td>
<td>-um</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>imię</td>
<td>name</td>
<td>-ię</td>
<td>yes</td>
<td>suppletion</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>zwierzę</td>
<td>animal</td>
<td>-ę</td>
<td>yes</td>
<td>suppletion</td>
<td></td>
</tr>
</tbody>
</table>
C Test Suite

C.1 Case Assignment

C.1.1 Verbs

1. The printer prints documents.
2. The printer does not print documents.
3. The printer prints pages, copies and documents.
4. The printer does not print pages, copies and documents.
5. You can scan documents.
6. You cannot scan documents.
7. Use these settings.
8. Do not use these settings.

C.1.2 Prepositions

9. an overview of software functions
10. text on the scanned item
11. documents that need to be processed with this device
12. adjusting the guides in the document feeder tray to the size of the item that you are loading
13. Inserting a pause or a wait into a fax number
14. Unplug the product telephone cord from the wall.

C.1.3 Nouns

15. printing a page
16. using the software

C.1.4 Relative Clauses

17. the copy which the printer prints
18. the copy which the printer does not print
19. the copy which the printer can print
20. the copy which the printer cannot print

C.2 Aspect

C.2.1 Main Clause

21. Print all pages.
22. Do not print all pages.
23. Keep in mind that the printer may not be working.
24. (When you are using the OCR software, follow these guidelines in order to obtain the
best possible results:) Process only pages that have crisp, clear text.
25. The printer is printing pages.
26. The printer was printing pages.
27. The printer has printed pages.
28. The printer has been printing pages.
29. The printer had been printing pages.
30. The printer will print pages.

C.2.2 Subordinate Clause
31. A fax is being received while the printer prints a job.
32. Jobs queue and wait until another job finishes.
33. Do not move the lever after the scanner has begun sending the page.
34. The software initiates the scan by using the special settings for the text.
35. In order to send long pages by using the printer, follow these steps:
36. (This product is specifically designed to allow you to do many tasks simultaneously.)
For example, you can print a document while you send a fax.
37. It may not be connected properly to the computer.

C.3 Word Order

C.3.1 Reversed Subject
38. There may be a poor phone connection.
39. You may also need to clean these parts if there are smudges or other marks on documents
or on scans of documents.

C.3.2 Verbal Adjuncts
40. The characters have been correctly interpreted by the OCR software.
41. (This user guide provides details about performing tasks from the product control
panel, changing infrequently used settings, resolving problems, and contacting
<trademark>HP</trademark> support.) This guide also provides an overview of software
functions.

C.3.3 Sentential Adjuncts
42. For example, if a fax is being received while you are printing a job, the fax is stored in
memory.
43. Alternatively, if the test page was too light, check whether the toner cartridge needs to
be replaced.
C.4 Anaphora Resolution

C.4.1 Object Pronouns

44. If you use the first procedure, then the setting remains on for all jobs until you change it again.
45. Use the two latches on the printer door to pull it open.
46. Locate the special media lever on the top edge of the control panel and move it to the right.
47. The device is receiving faxes but is not sending them.

C.4.2 Subject Pronouns

48. Text on the scanned item cannot be edited by a word-processor program unless it is first run through OCR software.
49. Fragile items can be damaged if they are unprotected.

C.4.3 Interclausal Anaphora

50. (Near the top, inside edge of the document release door, inspect the three pressure feet and the pad for dirt or debris.) If they are dirty, clean them with a lint-free cloth sprayed with a small amount of alcohol.

References


