# **Application Adaptive Electronic Dictionary with Intelligent Interface**

Svetlana Sheremetyeva

Copenhagen Business School, LanA Consulting Madvigs Alle, 9, 2 Copenhagen, Denmark, DK-1829 <u>lanaconsult@mail.dk</u>

#### Abstract

The paper presents an electronic dictionary that can be adapted to the needs of different NLP applications. It suggests some ways to save on software customisation and acquisition effort through an intelligent developer interface. The emphasis is made on the flexibility of data representation, handling and access speed.

#### 1 Introduction

In this paper we try to contribute to the problem of electronic dictionaries with a case study, -TransDict, - a multilingual lexicon for a family of patent-related NLP applications, such as AutoPat, APTrans and AutoRead<sup>1</sup>. TransDict thus conforms to the "Multilingual-Specialized" dictionary paradigm (Sérasset, 1993), but it can also be used as a stand alone tool and adapted for other language related tasks, e.g., training computational linguists.

The motivation to focuse on application tuned dictionaries is that though developing reusable fullsized knowledge bases for NLP systems is highly desirable this process is extremely expensive and time consuming, and reusability is not guaranteed. If an NLP system uses a restricted sublanguage, and, thus, can operate with smaller-scale dictionaries, the scope of acquisition and development effort will decrease correspondingly. Dictionary software should be adaptable to the specificity of sublanguages.

The languages that are currently covered are English and Danish but TransDict can easily be extended to a multiple number of other languages. TransDict features a powerful environment for acquisition, editing, browsing, defaulting and coherence checking. It is implemented in C++ as an integral part of 32-bit Windows applications for Windows 95/98/2000/NT.

# 2 Related work

A vast amount of research in the field of electronic dictionaries concentrate on data unification, representation, organization and management with the major focus on multilingual dictionaries as, for example, in (Wong, 2000; Boitet et al., 2002).

Multilingual electronic dictionaries often include a database of cross-referenced unilingual dictionaries with the use of interlingua such as ontology (Onyshkevich and Nirenburg, 1994)) or a pivotal language (Boitet et al.,cf.).

The architecture of such dictionaries normally include a lexical database and a set of tools for data management, - visualisers, editors, defaulters, etc. (Khatchadourian, 1992). A user-friendly interface is one of the major issues still uderdeveloped (Bilac and Zock, 2003).

XML and SGML data representation languages (Boitet et al., cf.) have been a successful approach to facilitate the export of electronic dictionaries to different applications though many dictionaries use their own internal data representation formats (Fedder, 1992).

Finally, it is desirable for electronic dictionaries to be stand-alone modules with defined interfaces for interaction with other linguistic applications (Pointer project report, http://www.computing.surrey.ac.uk/ai/pointer).

### **3** Overview of TransDict

#### 3.1 Feature space

TransDict is originally built over a set of features relevant for the patent applications including:

*Semantic features*: SEM\_Cl - semantic class, CASE\_ROLEs, - a set of case-roles associated with a lexeme, if any).

*Syntactic features*: FILLERs, - sets of most probable fillers of case-roles in terms of types of phrases and lexical preferences.

*Linking features*: PATTERNs, - linearization patterns of lexemes that code both the knowledge about co-occurrences of lexemes with their case-roles and the knowledge about their linear order.

<sup>&</sup>lt;sup>1</sup> AutoPat, APTrans, AutoRead, - computer systems for authoring, translation and improving readability of paten claims, correspondingly (Sheremetyeva, 2003)



Figure 1. An overall architecture of TransDict.

*Morphological features*: POS, - part of speech, MORPH, - wordforms, number, gender, etc.; the sets of parts of speech and wordforms are domain and application specific (Sheremetyeva, cf.).

*Rank feature*: RANK, - corpus-based frequency within one semantic class. The more frequent is a lexeme, the less its rank.

#### 3.2 Organization and architecture

TransDict includes cross-referenced monolingual lexicons for every language. A monolingual dictionary consists of a set of entries. An entry identifies lexical information for one meaning of a lexeme of a given language. Every entry is maximally defined as a tree of features:

## SEM-CL[Language[POS RANK [MORPH CASE\_ROLE FILLER PATTERN]

The CASE\_ROLE, FILLER and PATTERN features might not be specified in certain entries, e.g., for nouns-physical objects.

A maximal entry has the following fields: entry::= semantics SEM\_CL language LANGUAGE part of speech POS major-form string TAG other-forms {string TAG}+ case-frame {CASE\_ROLE}+ filler {CASE\_ROLE{FILLER}+}+ patterns {PATTERN}+ frequency RANK translation{cross-linguistic equivalent entry index}+

TAG is a label to code several features: POS, number, inflection type and semantic class: object, event, etc., providing for powerful tagging.

The architecture of TransDict is shown in Figure 1. All information is stored in TransDict internal formats: in data files and index files. The developer works with the Main Dictionary File (MDF) visualised by the interface (Figure 2).

Ad   Delete     English   SC     Congressional   O     Ad   Delete     Congressional   O     Add   Delete     Congressional   O     Add   Delete     Congressional   O     Add   Delete     Congressional   O     Add   Data     Congressional   O     Congressional   O     Congressional   O     Conside   D     Concel   Petrol Speech     Petrol Speech   Predicate passive     Code   PD     Conside   O     Conside   O  <	X				-		ct)	ary (TransDi		E AutoPat Transla
Add   Delete     English   SC POS     Language   English     Compessional   O ADJ     Commente   Forma     Connected   Poid     <								~		
English SC PoS   congressional 0 ADJ   congressional 0								Dalata I		
congressional   0   ADJ     congressional tem   0   ADJ     congressional tem   0   ADJ     congressional tem   0   ADJ     congressional tem   0   ADJ     congressional   0   ADJ     conside   0   <		ere you can type in comments	Comments: H	Code: C	on	n-class: Connection	Sem			
congressional ten   0   ADJ     conside   0   PO N     conside   0   ADJ     conside   0   ADJ     conside   0 <td></td> <td>eed language</td> <td>Comments: S</td> <td>Code: EN</td> <td></td> <td>auege: English</td> <td>Lan</td> <td>PoS 🔺</td> <td>SC</td> <td>English</td>		eed language	Comments: S	Code: EN		auege: English	Lan	PoS 🔺	SC	English
congressional   0   ADV     considence   PO   N     conicid   0   ADV     conicid   0   ADV     conicid   0   ADV     conicidence   PO   N     coniquid   0   ADV     coniquid   0   No     coniquid   0   ADV     connectid   C   Poi     connectid   C   ADV     cono				· · · · · · · · · · · · · · · · · · ·						
compressional   PO   N     compressional   O   ADJ     congressional   O   ADJ     congressional   O   ADJ     congressional   O   ADJ     congressional   O   ADJ     concid   O   ADJ     coning   O   ADJ     conjagetion   O   N     connectid   PO   N     connectid   PO   N     connectid   PO   N     connectid   PO   <		ere you can type in comments	Comments: H	Code: PD	bassive	peech: Predicate passiv	Part of Sp			
Congressional   O   ADJ     Concide   O   ADJ     Concidence   PO   N     Coning   G   ADJ     Coning   O   ADJ     Conjugate   O   N     Conjugate   O   N     Connected   Pdcn     Connectide   ADJ <t< td=""><td></td><td></td><td></td><td>vetlana</td><td>12/2003 11:49:01</td><td>monte: Added 16/12/20</td><td>Com</td><td></td><td></td><td></td></t<>				vetlana	12/2003 11:49:01	monte: Added 16/12/20	Com			
congressional   0   ADJ   English   Danish   Norwegian   Swedish   Russian   Case-frame     conjuent   0   ADJ   O   ADJ   O   Consider   Consider   Pdc   Pdc     conicidence   PD   N   Conscidence   Pdc   Pdc   Pdc   Ocosing   Conscidence   Pdc   Pdc   Pdc   Pdc   Pdc   Ocosing   Conscidence   Pdc   Pdc   Pdc   Pdc   Pdc   Pdc   Pdc   Ocosing   Conscidence   Pdc   P					12/2003 11:43:01	ments. [Added for targed	Com			
Conguence   PO   N     conguent   0   ADJ     concid   0   ADJ     coning   0   ADJ     conjugated   0   ADJ     connectable   0 <t< td=""><td></td><td></td><td>10</td><td>Russian</td><td>egian Swedish</td><td>Danish Norwegian</td><td>English</td><td></td><td></td><td></td></t<>			10	Russian	egian Swedish	Danish Norwegian	English			
congruent   0   ADJ     considence   0   ADJ     conicid   0   ADJ     conicid   0   ADJ     conicid   0   ADJ     conicid   0   ADJ     conicidence   PD   N     coning   0   ADJ     conjugate   PO   N     conjugating   0   FG     connected   Pdc   Pdc     connectable   0   ADJ     connectable   0   ADJ     connecting   0   ADJ     connecting   0   ADJ     connectable   0   AD	X	lase-frame					- Word forms -			
conic   0   ADJ     conical   0   ADJ     conical   0   ADJ     conical   0   ADJ     conical   Pdc     conical   Pdc     conical   Pdc     conical   Pdc     conical   Pdc     conical   Pdc     coning   0     ADJ   Connected     conigal   0     conjugate   0     conjugated   0     conjugating   0     connected   Pdcn     conjugating   0     connected   Pdcn     conjugating   0     connected   Pdcn     connected   Pdcn     connected   Pdcn     connecting   0     connecting   0 <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>			-							
conicidence PG N   conicidence PO N   conicidence PG N   coning O ADJ   coningal O ADJ   coningaling O PG   conjunction O N   connected Pdc   connected Pdc   connected Pdc   connected Pdc   connected Pd   connecting O	ame:	Choose case-roles for the current case-frame:				rms				
considence   PO   N     coning   Gamma   FG   N     coning   O   ADJ     coning   O   PG     coning   O   PG     coning   O   PG     coningdian   O   ADJ     conjugated   O   N     conjugation   O   N     conjunction   O   N     connectable   O   ADJ     connectable   O   ADJ     connection   O   N     connectine<			Pdc			nnected	GF cor			
coning   FG   N     coning   0   ADJ     coning   0   ADJ     coning   0   Pdcp     coningate   PO   N     conjugate   PO   N     conjugation   0   N     conjugation   0   N     conjugation   0   N     connected   Pdcp     conscipation   0   N     connected   Pdcp     connected   Pdcp     connected   Pdcp     connected   ADJ     connectide   0     connectide   0 </td <td></td> <td></td> <td>Pdcs</td> <td></td> <td></td> <td>connected</td> <td>C.Ss. ier</td> <td></td> <td></td> <td></td>			Pdcs			connected	C.Ss. ier			
coning   0   ADJ     coning   0   PG     coning   0   PG     coning   0   PG     coning   0   ADJ     coningated   0   PG     conjugation   0   N     connectable   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connecting   C   PG     connecting   C   ADJ     connecting   C   ADJ     connecting   C   ADJ     connecting   C   ADJ     connection   O   N     connection   O   N     connection   O   N     connection   O   N     connection   O   N<										
coning   0   PG     conjugate   P0   N     conjugation   0   PD     conjugation   0   N     connected   C   PO     connective   0   ADJ     connecting   C   PG     connective   0   ADJ     connectin			Pdcp			e connected	C Sp are			
conjugate   P0   N     conjugation   0   N     conjugation   0   N     conjugation   0   N     conjunctive   0   ADJ     connected   C   P0     connected   C   P0     connectible   0   ADJ     connecting   C   ADJ     connection   0   N     connection   0   N     connective   0   ADJ     connection   0   N     connective   0   ADJ     connective   0   N     connicisseu <td></td> <td></td> <td>Pdca</td> <td></td> <td></td> <td>ing connected</td> <td>C Abs bei</td> <td>PG</td> <td>Ō</td> <td></td>			Pdca			ing connected	C Abs bei	PG	Ō	
conjugal   0   ADJ     conjugated   0   PD     conjugated   0   PD     conjugated   0   PG     conjugated   0   PG     conjugated   0   N     conjugation   0   N     conjunction   0   N     conjunction   0   N     conjunction   0   ADJ     connectable   0   ADJ     connectible   0   ADJ     connecting   0   ADJ     connection   0   N     connection   0   N     connection   0   N     connection   0   N     connective   0   ADJ     connector   0   N     connector   0   N     connector   0   ADJ     connector <td></td> <td>means</td> <td>Delei</td> <td></td> <td></td> <td></td> <td></td> <td>N</td> <td>PO</td> <td></td>		means	Delei					N	PO	
Conjugated   0   PD     Conjugating   0   PG     conjugating   0   PG     conjugating   0   PG     conjugating   0   N     connective   0   ADJ     connecting   0   ADJ     connective   0   ADJ								ADJ	0	conjugal
conjugated   0   PG     conjugation   0   N     conjugation   0   N     conjugation   0   N     conjunctive   0   ADJ     connectable   0   ADJ     connectible   0   ADJ     connectible   0   ADJ     connecting   C   PG     connecting   C   PG     connecting   C   PG     connective   0   ADJ     connective   0   N     con			Pdcm		t l	ist be connected	C Sm mu	N		conjugate
conjugating   0   PG     conjugating   0   N     conjugating   0   ADJ     connectable   0   ADJ     connecting   C   PG     connecting   C   PG     connecting   C   PG     connective   0   ADJ     connectively   0   ADJ     connectively   0   ADJ     connectively   0   ADJ     connectively   0   N     connection   0   N   V <			Pdcc			n be connected	C Sc car			conjugated
conjunctive   0   N     conjunctive   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connectible   0   ADJ     connecting   C   PG     connectines   0   ADJ     connective   0   N     connective   0   N     connective   0   N     connective   0   N     <			1 400				1.00.00	PG		conjugating
conjunction   U   N     conjunction   0   ADJ     conjunctive   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connectable   0   ADJ     connectible   0   ADJ     connectible   0   ADJ     connecting   0   ADJ     connection   0   N     connection   0   N     connectively   0   ADJ     connector   0   N     connoiset										conjugation
connectable   O   N     connectable   0   ADJ     connectible   0   ADJ     connectible   0   ADJ     connecting   0   ADJ     connectioness   0   ADJ     connective   0   ADJ     connector   PO   N     cono							1			
connectable   0   ADJ     connectable   0   ADJ     connectible   0   ADJ     connectible   0   ADJ     connecting   C   PG     connections   0   ADJ     connections   0   N     connections   0   N     connective   0   ADJ     connective   0   N     connection   0   N     connective   0   N     connection   0   N     connot	. 1									
connected C PD   connectible 0 ADJ   connecting 0 ADJ   connection 0 N   connective 0 ADJ   connective 0 ADJ   connective 0 ADJ   connector PO N   connot	)   le	OK Cancel				ik): 1	Frequency (ran			
connectible   0   ADJ     connectible   0   ADJ     connecting   C   PG     connections   0   N     connections   0   N     connections   0   N     connective   0   ADJ     connective   0   N     connection   0   N     connoisseu   PO   N     connotation   0   N     connotatio				nle (suntax):	Cas					
Connecting   0   ADJ     connecting   Case-frame:   (2 indirect-obi)     Connecting   C   PG     connecting   C   PG     connection   0   N     connective   ADJ     connector   PO     connector   Itas 2,1     (1 x 3,2,4)						(d. 11. 0)				
Connecting   C   PG     connection   0   N     connectionless   ADJ     connectively   CON     connectively   PO     connectively   PO     connectively   PO     connectively   Next>>         Patterns:   [1 x 2)     [1 x 3 x 2]   [1 x 3 x 2]     [1 x 4]   [1 x 6 4]     [1 x 6 4]   [1 x 6 4]     [1 x 3]   [1 x 3]	-			( (N NF)	Sub		Case frame:			
connection   0   N     connectionless   0   ADJ     connective   0   N     connotiseseur   <				IN NPI	dire		Case-frame.			
Connectionless 0 ADJ   connective 0 ADJ   connector PO N   connexion 0 N   conniver PO N   connoiser PO N   connoiser<										
Connective   0   ADJ     connective   0   ADJ     connectively   0   ADJ     connectively   0   ADJ     connectively   0   ADV     connectively   0   ADV     connectively   0   N     (1 x 2)   (1 x 3 2 4)   (1 m tegnth" in neght" (in negones to" N) through one step"     (1 x 6)   (1 x 6)   (1 m tegnth" in tegnet to" NP[tyb" (entythrough" NP)     (1 x 6)   (1 x 6)   (1 m tegnet th" NP[tyb" (entythrough" NP)     (1 x 6)   (1 x 3)   (1 m tegnet th" NP] <td></td> <td>'to'' NP)(''of'' N)(''of'' NP) PP Inf)</td> <td>with" NP) ("to" N)("</td> <td>t-obj (["with" N) ("</td> <td>lindi</td> <td></td> <td>Uverwrite &gt;</td> <td></td> <td></td> <td></td>		'to'' NP)(''of'' N)(''of'' NP) PP Inf)	with" NP) ("to" N)("	t-obj (["with" N) ("	lindi		Uverwrite >			
connectively 0 ADV   connector P0 N   connector P0 N   connexion 0 N   conniver P0 N   conniver P0 N   connoiseur P0 N   connoisseur P0 N   connoiseur P0 N   (1 x 2) (1 x 24) (1 m leginet' ''in height''' (in response to'' NP)''on estep''   (1 x 64) (1 x 64) (1 m leginet'''in height''' (in response to'' NP)''through'' NP)   (1 x 64) (1 x 64) (1 m leginet'''in height''' (in response to'') ''through'' NP)   (1 x 3) (1 x 64) (1 m leginet'''in height''' (in response to'') ''through'' NP)   (1 x 3) (1 x 3) (1 m leginet'''in height''' (in response to'') ''through'' NP)										
connector   PO   N     connexion   0   N     conniver   PO   N     connoiseur   PO   N     connoistation   0   N      (1 x 2)   (1 x 24)     (1 x 3 x 2)   (1 x 42)   (1 x 3 x 2)     (1 x 6 4)   (1 x 6 4)   (1 x 6 4)     (x 1)   (1 x 3)   (1 x 3)		direction of" N)	on plane" ("in the	(PP "in a comm	plac					
connexion   0   N     connivance   P0   N     connivance   P0   N     connivance   P0   N     connoiseaur   P0   N     (1 x 2 4)   (1 x 2 4)   (1 x 4)     (1 x 6)   (1 x 6)   (1 x 6)     (1 x 6)   (1 x 6)   (1 x 6)     (1 x 1)   (1 x 3)   (1 x 3)			of" NP) "therein")	("in the direction						
connivance P0 N   conniver P0 N   conniver P0 N   conniser P0 N   connoiseur P0 N   connoiseur P0 N   connoiseur P0 N   connoiseur P0 N   (1 x 2) (1 x 3 x 2) (1 m leght/min leg		NDOL-SU NEDUSAKU	Calabaran Lacon - U		-					
conniciseur PO N   conniciseur PO						(1 + 2)				
connoisseur   PO   N     connoisseur   PO   N     connoisseur   O   N     connoisseur   O   N     ("In" N) PP "together" "In a fixed position" ("In the form of" N)     ("In" N) PP "together" "In a fixed position" ("In the form of" N)     ("In the form of NP("by") Gettight" Gettight" ("In the form of" NP)     (*   In a naje!" ("In" Adj "location") ("In" do chact with" N)     (*   Next>>     (In the form of NP) ("In the spect to" N)     (In the form of NP) ("In the spect to" N)     (In the form of NP) ("In the spect to" N)     (In the form of NP)     (In the form of NP) </td <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Patterns:</td> <td></td> <td></td> <td></td>							Patterns:			
connotation     N     (1 3 x 2 4)     ("in the form of" NP]"by" Gerg["through" NP]" NP)       (     I x 6)     "at an angle" ("in" Adj "direction"] ("into contact with" N)       (     I x 6)     "at an angle" ("in" Adj "direction"] ("into contact with" N)       (     I x 6 4)     "into contact with" NP]" (without" NI" without" NP)       (Into contact with" NP]" (without" NI" without" NI" without" NP)     Into contact with" NP]       (Into contact with" NP]" (without" NI" without" NI")     Into contact with" NP]										
(1 x 6)     "at an angle" ("in" Adj "direction") ("into contact with" N)       (<										
(X + 6.4)     ("into contact with" NP)("with respect to" NP)       (X + 1)     ("into contact with" NP)("without" NP)       (X + 3)     ("into contact with" NP)		'into contact with" N)	("Âdj <sup>(</sup> direction")("	"at an angle" ("ir		(1 x 6)				•
(1 x 3) ("in accordance with" N) ("in accordance with" NP)										
								Next >>		<< Previous
Shown: 336 Total: 41303	-	Ill'according to" N	"If"in" Adi "mode"	("in" adi "positior		[( × 3)			al: 41309	Shown: 998 Total
😫 Start 🛛 🔏 🛱 🔤 🔹 🕞 APTmultApr04_MishaMarch 🔚 Transdict 🖓 Untitled - Paint 🕢 🖓 👘 🤤 🗐 🐺 🍕	N 🛃 🍝 6:41 PM		Paint		sdict	aMarch	TmultApr04_Mish	» 🖂 🖂 AP	) 🖾 📾	🛃 Start 🔢 🚮 🧔

Figure 2. A screenshot of the TransDict interface displaying the entry for the lexeme "connected"

When the lexicographer saves the data multiple extractions from MDF are automatically created. These extractions contain different data subsets relevant for different processing steps (tagging, disambiguation, transfer and generation). The extractions are created for every language and for every pair of languages. They are linked to applications by special DLL (dynamic link library) functions that access only one of the dictionary extractions for every processing step. This approach gives a significant increase in access speed and processing, which is crucial for real world systems. This and the fact that TransDict is implemented for PC motivated our choice not to use the SOL database and XML (which would have slowed down the application performance). It does not mean, however, that TransDict could not be used in the on-line regime. An interface and a dll can be written for this purpose.

# 4 Supporting tools

We developed the following TransDict tools: *Data importer/merger* imports wordlists and/or feature values from external files and applications. For example, the tool is pipelined to a tagger and to AutoPat and AutoTrans user interfaces, to automatically import unknown words.

*Defaulter* automatically assignes entry structures and some of feature values to entries.

*Editor* a) edits feature values in an entry and b) edits dictionary settings, - languages, semantic classes, parts of speech, wordforms and their tags. Any change of settings automatically propagates to corresponding entries.

*Morphological generator* automatically generates wordforms for a given word base form.

*Content and format checker* reveals incomplete and/or bad formatted entries.

*Look-up tool* performs wild card search and search on any combination of specified parameters.

# 5 Interface design

A lexicographer interacts with the lexicon by an extemely user-friendly interface (Figure 2). The left pane of the interface screen contains a scrollable list of lexeme base forms<sup>2</sup> in a selected language. A click on a language bookmark over

<sup>&</sup>lt;sup>2</sup> For convenience other wordforms are not included in this list but can be displayed on mouse click.

	Sem-class:	Physical-object	Code: PO	Cor
	Language:	English	Code: EN	Cor
E Part	of Speech:	Noun	Code: N	Cor
	Comments:	from old dictionary		
English	Danish	Norwegian Sw	vedish Russian	
- Word for	ns —		896 D.S.S.	
Major	Forms			Tags
ΘS	device			N
O PI	devices			Np

	Sem-class:	<sup>o</sup> hysical-object	Code: PO	Cor
	Language:	Danish	Code: DA	Cor
Part	of Speech:	VounCommon	Code: N	Cor
	Comments: f	rom old dictionary		
English	Danish	Norwegian Swed	lish Russian	
English Word forr		Norwegian Swed	lish   Russian	
		Norwegian Swed	lish   Russian	Tags
Word for	ms	Norwegian Swee	lish   Russian	Tags N
Word for Major	ms Forms	, - ,	lish   Russian	
Word forr Major I S	ns Forms indretning	ər	lish Russian	N

Figure 3. A fragment of English and Danish equivalent entries as shown in the interface.

the morphological zone displays an entry in the selected language equivalent to a highlighted word in the left column. All supporting tools are accessed through the interface menus.

The "Add" button calls pop-up menus where the developer is prompted to select a semantic class and part-of speech. This done, an entry with a relevant structure, tags and default values will be displayed. After the user types in a base form all other wordforms are automatically generated on mouse click. The developer is to review the default knowledge and edit it if necessary. The content and format checker take care of correct descriptions with different kinds of alert messages and rewriting support. Powerful search can be done both in a look-up and edit mode.

Changing the dictionary settings can easily change a base form status of a wordform, the structure of the entry and other specification parameters. Figure 3 shows how the default noun entry with two slots for its morphological forms: singular and plural, is reset for Danish where definiteness is expressed morphologically, thus duplicating the number of members of the noun paradigm compared with English.

#### 6 Conclusion

In this paper we described an on-going project on developing a multilingual electronic dictionary, - TransDict, integrated with patent domain applications. We focused on such effort saving strategies as knowledge organization, access, reusability, support tools and interface design. As of now (April 2004) the dictionary program including intelligent application adaptive interface integrated with supporting tools and external applications, - AutoPat, AutoTrans, AutoRead (Sheremetyeva, cf.) is fully implemented and tested. This "shell" can now be used to create any number of dictionaries with different feature spaces.

The TransDict patent domain knowledge base currently contains about 60,000 completed English entries and around 100 equivalent Danish entries that are directly used in testing analysis, transfer and generation modules for the English-Danish machine translation system. We plan to increase the English-Danish knowledge base to a product size level by December 2004.

TransDict (with patent domain or other knowledge) can be used as a stand-alone tool, for other applications e.g., for training computational linguists.

#### References

- S.Bilac and M.Zock. 2003. *Towards a user-friendly dictionary interface*. Papillon 2003 Workshop, 3-5 July, NII, Sapporo, Japan.
- C.Boitet, M.Mangeot-Lerebours and G.Sérasset. 2002. The PAPILLON project: cooperatively building a multilingual lexical data-base to derive open source dictionaries & lexicons. Proceedings of the 2nd Workshop NLPXML 2002, Post COLING 2002 Workshop. Taipei.
- L.Fedder.1992, *The Multilex Internal Format*. Multilex report, June.
- H. Khatchadourian 1992, *Tools, functional specifications*. Multilex report, February.
- B. Onyshkevich and S. Nirenburg. 1994. *The lexicon in the scheme of KBMT things*. Thechnical report MCCS-94-277, CRL, NMSU.
- G. Sérasset. 1993. Recent Trends of Electronic Dictionary Research and Development in Europe, Technical Memorandum Electronic Dictionary Research (EDR), Tokyo, Japan.
- S. Sheremetyeva. 2003. Natural Language Analysis of Patent Claims. Proceedings of the workshop "Patent Corpus Processing" in conjunction with 41<sup>st</sup> Annual Meeting of the Association for Computational Linguistics (ACL 2003), Sapporo. Japan, July 7-12.
- K.Wong.2000. Multilingual Electronic Dictionary Project.http://www.csse.monash.edu.au/hons/pr ojects/2000/Kevin.Wong/ksgw.htm