Representing natural gender in multilingual lexical databases*

Noam Ordan and Shuly Wintner
Department of Computer Science, University of Haifa
shuly@cs.haifa.ac.il

Abstract

Natural languages encode gender distinctions in various ways. We investigate the differences between English and Hebrew in this respect, our departure point being the relations that are defined between the feminine and the masculine realizations of nouns in the English WordNet. We define a number of distinct classes of English nouns which differ in the way they realize gender distinctions. We then define similar classes of Hebrew nouns and show how to map the Hebrew nouns (and relations defined over them) to the English structure. This establishes a systematic assignment of Hebrew nouns to WordNet synsets, which is consistent with the ideas underlying multilingual extensions of WordNet. The main result is a consistent Hebrew WordNet which is aligned with the English one, but an additional contribution is a set of desiderata for the correct encoding of (systematic) semantic differences among languages.

1 Introduction

Languages differ in the ways they encode gender (Corbett 1991; Hellinger and Bußman 2001). For our purposes here, the main observation is that some languages (e.g., French, German, Italian or Hebrew) have *grammatical* (or, more specifically, *morphological*) gender, while others (e.g., Turkish, Finnish or English) do not. But even languages which do not morphologically mark gender on nouns often distinguish between nouns denoting masculine and feminine entities. Such distinctions are referred to as *natural* gender and can be manifested in various ways. English, for example, realizes natural gender distinctions both lexically (through the use of different forms for nouns denoting masculine and feminine entities) and grammatically (through gender agreement in pronouns). In contrast, languages with grammatical gender usually encode natural gender morphologically, via inflectional affixes. In this work we focus on nouns denoting entities with biological gender (humans, human groups such as professions, animals, etc.), for which we use the term 'animate nouns' in the sequel. We are mostly concerned in this work in *natural* gender, and

*To appear in the International Journal of Lexicography 18:3, September 2005.

future reference to morphological gender will only be made (in reference to Hebrew) in order to emphasize the discrepancies between the two concepts.

The relationship between word senses among different languages is not always a one-to-one relationship. A word sense in L_1 can be matched to anything in L_2, from a *gap* (zero lexical correspondence between the languages), through a partial match (two or more words are matched from L_1 to L_2 or vice versa) to a full match (a one-to-one lexical correspondence). In the usual case, there is no regularity of meaning assignment to certain words; the cross-lingual matching cannot be predicted, and creating a cross-lingual database is a matter of extensive, non-systematic manual work of lexicographers. However, since animate nouns behave in a relatively systematic way in languages, a more general solution can be devised.

Our main objective is to find a consistent mapping of Hebrew nouns to synsets of the English WordNet (Fellbaum 1998), a lexical-semantic database; this is in accord with a methodology for developing multilingual extensions of WordNet advocated by the MultiWordNet paradigm (Bentivogli, Pianta, and Girardi 2002). We first (section 2.1) classify English animate nouns according to how gender distinctions are realized, and define six distinct classes of such nouns. In section 2.2 we shift our attention to Hebrew, a language with morphological gender. We show that Hebrew animate nouns can also be classified along similar dimensions, although different classes are induced. In section 3 we describe the ways in which animate nouns in the different classes are represented in WordNet and point at some inconsistencies. We investigate alternative approaches to establishing a mapping of Hebrew animate nouns to the English WordNet in section 4 and define a set of desiderata for such a mapping. These desiderata can be useful for the correct encoding of (systematic) semantic differences among languages. The main result, however, is a structural organization of nouns in a multilingual lexical database in a consistent way which guarantees the usability of the system both as a standalone monolingual WordNet and for multilingual applications. We discuss the advantages of our proposed solution in section 5.

2 The linguistic manifestation of natural gender

2.1 Natural gender in English

English does not have morphological gender (Hellinger 2001). However, certain nouns which denote animates have distinct forms for the masculine and the feminine. We classify the animate entities of English according to how gender distinctions are realized. The criteria are whether distinctly masculine, distinctly feminine, or gender neutral forms (denoting the same entity) exist. Note that it is not always easy to determine whether a particular noun is gender neutral or masculine.[1] We rely on dictionary definitions (including WordNet) in order to make such decisions.

To define the classes, we selected 120 English nouns denoting animate entities; most were chosen manually, but to obtain better coverage we also used WordNet (version 1.6) and extracted nouns which are likely to be animate by looking at hyponyms of the *person* and *animal* synsets

[1] For example, *One of my favorite authors is Virginia Wolf* is probably better than *One of my favorite actors is Julia Roberts*, indicating that *author* is more likely to be gender neutral, whereas *actor* is probably dominantly masculine.

(note that there are 6531 nouns in the first set and 3997 in the second, so an exhaustive characterization of animate nouns in English is a matter of a much larger-scale research than ours here). We observe six classes (where N, F, M stand for *neuter, feminine* and *masculine*, respectively):

N This is the English default class: nouns in this class denote both males and females. Examples include *citizen, elephant, engineer, expert, messenger, neighbor, teacher.*

F Nouns which refer to females and have no masculine counterparts: *babyminder, bellydancer, callgirl, co-ed, concubine, first lady, housewife, midwife, showgirl, slavey, streetwalker.* Of the 120 animate nouns we have considered, 34 of the nouns are in this class.[2]

M Nouns which denote males and have no feminine counterparts: *dandy, gentleman, hunk, stud, womanizer.* We collected 13 nouns in this class.

MFN Nouns in this class have three distinct forms, denoting males, females and gender unspecified entities. Examples include *barman/barmaid/bartender, boy/girl/child, son/daughter/child, brother/sister/sibling, father/mother/parent, king/queen/monarch.* This class has 23 members in our study.

MF This class includes nouns which have distinct masculine and feminine forms, but no gender neutral one. Examples: *boyfriend/girlfriend, lord/lady, male/female, prince/princess, uncle/aunt.* 17 nouns are in this class.

FN This class includes nouns which have two distinct forms, one denoting females, the other gender neutral. Examples include *actor/actress, author/authoress, aviator/aviatrix, heir/heiress, hero/heroine, gay/lesbian, lion/lioness, usher/usherette.* This is by far the largest class (next to the default one), with 39 members in this study.

Note that one more class is logically possible, namely nouns with distinct masculine and gender neutral forms, but no feminine form. We found no such cases in English.

2.2 Natural gender in Hebrew

Hebrew has morphological gender, and hence gender distinctions are marked in more cases than English (Tobin 2001). In particular, most Hebrew nouns which denote humans or animates inflect for gender and have both masculine and feminine (but no neuter) variants (there are three different feminine suffixes in Hebrew: ◌ָה *"ah"*, ◌ֶת *"et"* and ◌ִית *"it"*). Hebrew nouns which denote non-animate entities are either morphologically masculine or morphologically feminine; animate nouns usually have two distinct forms, and morphological gender usually coincides with natural gender.

However, this is not always the case. The following discussion classifies Hebrew animate nouns into semantic classes according to how *natural* gender is expressed lexically and morphologically (here, H stands for *Hebrew*, whereas N, F, M are as above).

[2]Many of the nouns in this class have derogatory meanings. There are also some shared semantic characteristics to most nouns in class M. We defer a sociolinguistic investigation of this phenomenon to future research.

HMF The default class includes nouns which exhibit both forms. Examples include אזרח/אזרחית "*citizen*", סופר/סופרת "*author*", מוֹרֶה/מוֹרָה "*teacher*", חמור/אתון "*donkey*", פיל/פילה "*elephant*", נסיך/נסיכה "*prince*", ילד/ילדה "*child*", אריה/לביאה "*lion*", גמל/נאקה "*camel*", עבד/שפחה "*slave*". Note that while most nouns in this class use morphological suffixes to denote gender differences, in some cases two unrelated lexical entries exist for the masculine and the feminine.

HF Some nouns occur only in the feminine, have no masculine form and denote only females: שיננית "*dental hygienist*", מינקת "*wet nurse*", זונה "*whore*", מבשלת "*cook*", פיה "*fairy*". Arguably, the nouns גננת "*kindergarden teacher*" and מזכירה "*secretary, administrative assistant*" are in this class, since their masculine counterparts denote different entities (*gardener* and *secretary, officer of state*, respectively). Thus, to denote a male dental hygienist, for example, one would use *male dental hygienist*. As another example, fairies are by default feminine in Hebrew, unless explicitly referred to as *male fairy*.

HM Similarly, some nouns occur only in the masculine (i.e., cannot be morphologically inflected as feminine) and denote only male entities: כמר "*priest*", שטן "*devil*", גיגולו "*gigolo*".[3]

HMFN This class includes entities for which three forms exist, denoting masculine, feminine and neutral: אב/אם/הורה "*father/mother/parent*".

HN Certain nouns, mostly borrowed or acronyms, are gender-neutral and are not usually inflected for gender in Hebrew. These include עוּבָּר "*fetus*", אדם "*person*", דוקטור "*doctor*", יור "*chairman*", דינוזאור "*dinosaur*", קולגה "*colleague*", פרימדונה "*prima donna*", סלבריטי "*celebrity*", פנומן "*phenomenal person*", חלאה "*scum*", נשמה "*darling*", מפלצת "*monster*" and a few others. Note that it is possible to inflect some of these nouns morphologically by attaching a feminine suffix (by default, ית "*it*"). While such forms may be unaccepted today, they may become more frequent in the future and such nouns will hence shift to the default class. Note also that the last three examples are actually feminine nouns grammatically, but are used to denote humans of both sexes, so that the following are perfectly grammatical: הוא נשמה טובָה "*he's a kind*-FEM *soul*-FEM"; הוא חלאה אמיתית "*he's a real*-FEM *scum*-FEM". In contrast, the first two are masculine grammatically, but are used to denote both sexes: היא אדם חביב מאוד "*she's a very nice*-MASC *person*-MASC".

To these one must add also nouns which are morphologically either masculine or feminine, but semantically denote entities of both sexes, usually "lower" forms of animals, presumably those whose sex it is unimportant to specify. Feminine examples include נמלה "*ant*", תולעת "*worm*", לטאה "*lizard*", יונה "*pigeon*"; masculine examples include לויתן "*whale*", נחש "*snake*" and צפרדע "*frog*".

[3]The sociolinguistic reasons for the fact that such nouns are realized in one of the genders only are outside the scope of this work.

3 Natural gender in WordNet

3.1 Multilingual lexical databases

WordNet (Fellbaum 1998) is a network of words (nouns, verbs, adjectives and adverbs) over which several relations are defined. Words are organized as synonym sets, referred to as *synsets*, the idea being that several words are in the same synset if they all share one synonymous sense. For example, the word *horse* is assigned to six different synsets (implying that it has six distinct senses). One of these synsets includes also the word *knight*, and of course the sense of both *horse* and *knight* in this synset is the chess sense.

For our purposes here it is only important to recall that WordNet defines the following relations over synsets:

Hypernyms/hyponyms relate a noun which denotes a set of entities to a noun whose denotation is a superset/subset of that set. For example, *dog* is a hyponym of *canine* and a hypernym of *dalmatian, lapdog, mutt* and *poodle*, among others. A synset can have more than one hypernym and zero or more hyponyms.

The hypernym relation defines a hierarchy of synsets which is central to the organization of WordNet. In this hierarchy, *dog* is a hyponym of *canine* which is itself a hyponym of *carnivore*, going through *mammal* and *vertebrate* and *animal* all the way to *entity*. We sometimes refer to the hypernym relation as *direct* or *immediate* hypernymity, to distinguish it from its transitive closure relation.

Coordinate terms Two synsets are coordinate terms if they are both direct hyponyms of the same sysnet. For example, the coordinate terms of *dog* include *bitch, fox, jackal* and *wolf* because they are all immediate hyponyms of *canine*.

Antonyms are sometimes defined over nouns in WordNet, although their main use is for adjectives (Miller 1990). Interestingly, antonymy frequently holds for two nouns differing only in gender, as in *uncle/aunt* or *male/female*, although other types of nouns are also defined as antonyms (e.g., *victory/defeat*). The psycholinguistic indication for antonyms is taken to be the informal definition 'two words are antonyms if each is given on a word association test as a most common response to the other' (Miller 1990). This may explain *male/female* but not necessarily *duke/duchess*. Consequently, the definition of antonymy over gender-distinct nouns in WordNet is somewhat inconsistent.

Following the success of WordNet, similar networks for other languages have been developed. Our work is cast in the paradigm of MultiWordNet (Bentivogli, Pianta, and Girardi 2002) which strives to develop networks for other languages which are aligned with the Princeton English WordNet. This implies that word senses in other languages are mapped to existing WordNet synsets, thereby preserving as much as possible existing relations over synsets. The underlying assumption is, of course, that synsets represent, in many cases, language independent notions, and thus relations over them should be transferable across languages.

One of the main issues involved in adding a new language to MultiWordNet is the extent to which the structure of the English WordNet should be retained. As we show below, the organization of animate nouns in WordNet is sometimes inconsistent, and in particular inadequate for Hebrew, in which gender distinctions are far more productive than in English.

3.2 WordNet representation of gender-sensitive nouns

In this work we would like to use additional perspectives on gender, especially a semantic one, namely, the structural organization of gender-sensitive nouns in a language's lexicon. This comes in accordance with our initial motivation, inspired by the MultiWordNet project, to offer a general solution for representing animate nouns of different languages in a single shared database, adhering to the double intention of keeping the "true" internal organization of these words in each of the languages and representing the different lexicons in relation to each other at the same time. Since it is assumed that synsets and the relations defined over them are largely language independent, it also ensures inheritance of semantic relations.

WordNet employs various strategies to encode nouns denoting entities with natural gender. We now explore some of these strategies, focusing on the various WordNet relations defined between gender-distinct nouns. As the following discussion shows, there is some inconsistency in the structural encoding of animate nouns in WordNet. We consider only nouns for which more than one form exists; trivially, nouns in classes N, F and M are represented as a single node in WordNet and are therefore excluded from the following investigation.

When more than one distinct form exist for the realizations of an animate entity, five different strategies are employed (synsets are given as lists of representative synonyms below):

One synset The masculine and the feminine (and, sometimes, also the gender neutral noun) are in the same synset. Examples include *cyborg, bionic_man, bionic_woman*; *freedman, freedwoman*; *juror, juryman, jurywoman*; *clansman, clanswoman, clan_member*; *president, chairman, chairwoman, chair, chairperson*; *sport, sportsman, sportswoman*; and *yachtsman, yachtswoman.*

Coordinate terms The masculine and the feminine are coordinate terms, direct hyponyms of the gender neutral form, if such exists, or of an "artificial" (i.e., non-lexical) synset. For example, *lesbian* and *gay_man* are hyponyms of *homosexual, homo, gay*; *granddaughter* and *grandson* are hyponyms of *grandchild*; *spokeswoman* and *spokesman* of *spokesperson*; and *bondwoman* and *bondman* of *slave.* Examples with an artificial hypernym include *merman, mermaid* which are hyponyms of *imaginary being*; or *Englishman, Englishwoman* which are hyponyms of *English_person.* Interestingly, there are also "reverse" examples, where artificial (non-lexical) hyponyms are added: *aristocrat* has as its hyponyms *male_aristocrat* and *female_aristocrat*; *oldster* has *old_man* and *old_woman* as hyponyms. Note that in some cases the coordinate terms are feminine and gender neutral (rather than masculine), as in *magician* and *enchantress*, whose direct hypernym is *occultist.*

Indirect coordinate terms While many times the masculine and the feminine have a common *direct* hypernym, sometimes their least common ancestor in the hypernym hierarchy is far-

ther up the tree. Examples include *black_woman* and *black_man*; *Mr* and *Mrs*; *girlfriend* and *boyfriend*; and *girl* and *boy*; all have the same least common hypernym, the sysnset of *person.* Similarly, *count* and *countess* are indirect hyponyms of *aristocrat* and *daughter* and *son* are indirect hyponyms of *child.* Note that the masculine and the feminine forms are not always symmetric, in the sense that sometimes different paths lead from the two to their least common ancestor.

Hyponym The most common WordNet strategy, encoding many of the nouns in class FN, is to define the female-denoting noun as a hyponym of the gender-neutral noun.[4] Thus, for example, *aviatrix* is a hyponym of *aviator, Scotswoman* of *Scot* and *usherette* of *usher.* This structure is particularly common when the female-denoting noun is morphologically derived from the gender-neutral one. It is also common in some cases of nouns in class MF, especially when it is hard to determine whether the non-feminine noun is indeed masculine or gender neutral: *actress* is a hyponym of *actor* and *cowgirl* of *cowboy*, for example.

Chain In one extreme case we found a chain organization: *businessperson* is the hypernym of *businessman* which in turn is the hypernym of *businesswoman.*

In parallel to the hypernymity organization, sometimes the masculine and the feminine are defined also as antonyms. This is the case for *son/daughter, lord/lady* and *male/female*, for example.

Table 1 summarizes the various WordNet organizations of gender-sensitive nouns, listed according to the classification of section 2.1. The columns of the table refer to the classes of section 2.1, and the rows to the WordNet organization. The entries specify the number of nouns in each combination of class and organization.

class:	MFN	MF	FN
One synset	5	3	0
Coordinate terms	17	2	2
Indirect coordinate terms	1	11	0
Hyponym	0	1	37
Chain	1	0	0

Table 1: WordNet organization of animate nouns

4 Representing gender distinctions in multilingual lexical databases

Our main motivation in this work is to define a consistent strategy for mapping Hebrew words to existing English synsets. Such a mapping should, on the one hand, be true to the phenomena of

[4] A thorough investigation of the cases in which the feminine noun is considered a hyponym of the masculine or gender neutral noun may be interesting from a sociolinguistic point of view, but is outside the scope of our work.

the Hebrew language and, on the other hand, comply as much as possible with the structure of the English WordNet. Obviously, these two requirements are in many cases contradictory.

The MultiWordNet framework provides a useful mechanism for reconciling structural differences between languages. When a word sense in one language cannot be mapped to an existing synset, a *lexical gap* can be defined (Bentivogli, Pianta, and Pianesi 2000; Bentivogli and Pianta 2000). Lexical gaps are originally defined as a situation in which 'a language expresses a concept with a lexical unit whereas the other language expresses the same concept with a free combination of words'. While it is unclear whether the gender discrepancies referred to above meet this definition, clearly the mechanism of coping with lexical gaps suggested by Bentivogli and Pianta (2000) is suitable in our case.

Janssen (2004) suggests several ways to deal with lexical gaps in the construction of multilingual lexical databases. We will illustrate the various approaches through the default gender distinctions in English and Hebrew, using as example the words אזרח/אזרחית "*citizen*". First, Janssen (2004) distinguishes between the *project-down* approach and the *hypernimity* approach. In the former, the word sense of the hypernymic word is discarded and is replaced by more specific meanings. In our example, the synset of *citizen* would have to be replaced by two more specific synsets, for *male citizen* and *female citizen.* As Janssen (2004) points out, this is problematic for many reasons, mainly because it introduces an artificial ambiguity in English and because with more languages added to the database, the number of synsets is likely to explode.

The hypernymity approach, on the other hand, explicitly models *citizen* as a hypernym of both its Hebrew translations. There are three variants to this approach, depending on whether this modeling is done without an interlingua, with an unstructured interlingua or with a structured one. The option of an unstructured interlingua is advocated by the EuroWordNet project (Vossen 2004), and its disadvantages are discussed by Janssen (2004). MultiWordNet, in contrast, is an example of a multilingual lexical database system which employs a structured interlingua: word senses in a number of languages are mapped to language-independent synsets, and while synsets are originally defined as in the Princeton English WordNet, more can be added to the system where needed. Crucially, structural relations, and in particular hypernymity, are defined over synsets and not over words. This facilitates the organization depicted in Figure 1, exemplified on the *citizen* case, where solid arrows indicate the hypernym relation and dashed lines map words to synsets, and where synsets are depicted using a representative member, typeset in small capitals and enclosed in an oval.

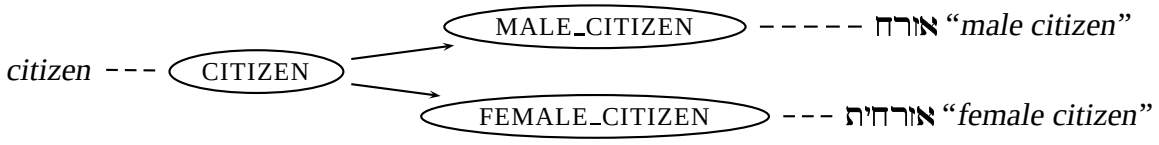

Figure 1: The structural organization of animate nouns, default case (classes N, HMF)

More generally, the solution we advocate is to introduce a synset for the masculine, feminine and gender-neutral variants of each animate noun, and define the masculine and feminine synsets as hyponyms of the gender-neutral one. In the default situation, exemplified by the structure of Fig-

ure 1, Hebrew masculine and feminine forms are assigned to the masculine and feminine synsets, respectively, whereas the English gender-neutral form is mapped to the gender-neutral synset.

Furthermore, this general solution immediately accounts for the non-default classes of animate nouns in both English and Hebrew, discussed in section 2. Thus, English nouns in class F and Hebrew nouns in class HF, for example, which denote females only, will be mapped to the feminine synset; but masculine and gender-neutral synsets for such nouns will nonetheless be present in the system, in case some other language realizes them. The Hebrew class HF is exemplified in Figure 2, matched against an English counterpart of class N.

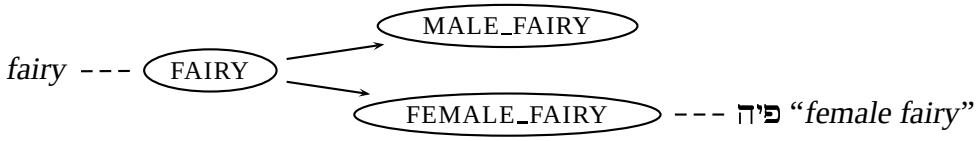

Figure 2: The structural organization of animate nouns of class HF (matched against a noun of class N)

The interesting case is that of the substantially populated class FN, where a feminine form and a gender-neutral form both exist in English, and where WordNet defines the former as a hyponym of the latter. These cases are organized following the example in Figure 3. Compare this organization with Figure 4 which exemplifies the situation of both masculine and feminine nouns where no gender-neutral form exists (class MF).

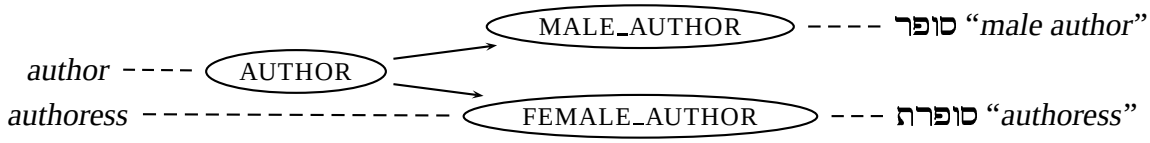

Figure 3: The structural organization of animate nouns of class FN

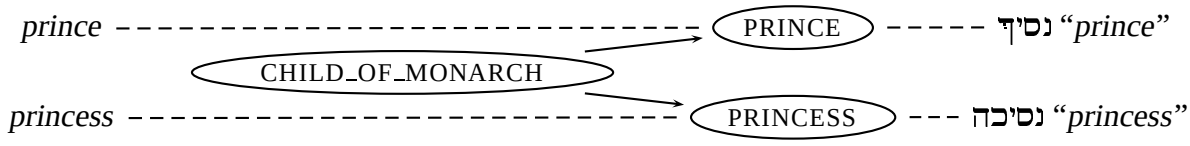

Figure 4: The structural organization of animate nouns of class MF

Finally, our proposed solution naturally accounts for animate nouns of class MFN, where three distinct forms exist to denote masculine, feminine and gender-neutral nouns. This is illustrated in Figure 5.

In way of conclusion, we propose to treat *all* the animate nouns of each language in a multilingual lexical database uniformly (compare to the current non-uniform representation of WordNet, as illustrated in table 1). Our solution calls for associating three distinct synsets with each animate entity, for the female, male and neuter possible realizations of the entity in any natural language.

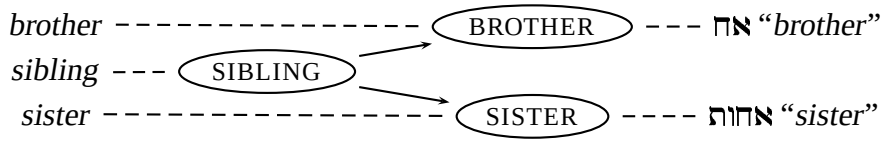

Figure 5: The structural organization of animate nouns of class MFN

5 Discussion

We proposed a general and consistent solution to the problem of representing the various ways in which English and Hebrew encode gender differences in animate nouns. The solution calls for the introduction of specific synsets for the masculine, feminine and gender-neutral variants of each animate noun, independently of whether or not these concepts are realized as lexical items in any particular language. The gender-neutral synset is to be a hypernym of both the masculine and the feminine synsets.

Furthermore, we propose to add two special synsets, namely for *male* and *female*, and to relate all the MALE_X synsets to MALE by hyponym relations; and similarly for the FEMALE_X synsets. In other words, a synset such as MALE_CITIZEN is viewed as a hyponym both of CITIZEN and of MALE.

The advantages of the proposed solution are manifold. First and foremost, it is consistent with the linguistic data. No artificial ambiguity is introduced explicitly or implicitly, as each noun form is mapped to an internal representation (a synset) which is true to the noun's interpretation. Furthermore, the internal representation is dependent only on the semantics and not on the linguistic realization of the semantic concepts in any particular language. Thus, the fact that Hebrew has no feminine form for the noun נחש "*snake*" is viewed as incidental; it does not rule out the possibility that in other languages a feminine form of *snake* may be realized lexically, and indeed a synset for *female snake* is part of the network.

The ability to account for more languages is yet another advantage of our solution. Some languages are said to have more than three morphological genders, but we claim that these cases should be viewed as noun classes, as in actuality they have little to do with biological gender. Those languages which encode natural gender using linguistic means, whether lexically or morphologically, can be simply and naturally added to the MultiWordNet paradigm in the same way as English and Hebrew.

An additional benefit is the systematic encoding of gender distinct nouns in English. As we have shown in section 3.2, WordNet currently is inconsistent in the representation of such nouns. For example, *grandson* and *granddaughter* are coordinate terms (both hyponyms of *grandchild*); but *son* and *daughter* are not, the former being a hyponym of *male_offspring* and the latter of *female_offspring*. Our solution adds a level of consistency to the network.

Finally, other relations which may be defined over the synsets can be carried over to all the languages represented in the system in a direct way. In particular, if *son* and *daughter* are antonyms in English, it is possible to extend the same lexical relation over to their Hebrew translation equivalents. Note that a lexical relation such as antonymy does not extend across languages automatically,

and it is up to the designer of the system to decide whether or not to extend it. However, semantic relations such as meronymy do carry over from one language to another.

One disadvantage of the proposed solution is that it requires many more synsets to be added to the system. This goes against the spirit of the MultiWordNet paradigm, which attempts to retain the original Princeton WordNet wherever possible. However, we maintain that such an extension is unavoidable. As more languages are added to the system, distinct masculine and feminine lexical items would be more likely to be realized lexically, and such lexical distinctions will inevitably require more nodes in the network.

Another possible criticism has to do with the fact that at least in Hebrew, the vast majority of animate nouns have feminine forms which are morphological inflections of the unmarked masculine form. It would seem advantageous to store only the masculine Hebrew nouns in the system and generate the feminine forms upon demand. However, such a solution would require storing, for each and every animate noun, at least two pieces of information: that it is animate; and what its feminine suffix is, since Hebrew has three different feminine nominal suffixes and the choice of suffix is mostly lexical. We maintain that it would be just as general to store the complete feminine form itself, with a synset as its interpretation.

Acknowledgements

We are grateful to Emanuele Pianta and Iris Eyal for their continuous support and useful comments. Thanks are due also to Lusia Bentivogli, Sara Kaufman, Nurit Melnik and Danny Shacham. This research was funded by the Israeli Ministry of Science and Technology, under the auspices of the Knowledge Center for Hebrew Telecommunication, with additional support from the Caesarea Rothschild Institute for Interdisciplinary Applications of Computer Science at the University of Haifa.

References

Bentivogli, L. and E. Pianta (2000). Looking for lexical gaps. In *Proceedings of Euralex-2000 International Congress*, Stuttgart, Germany.

Bentivogli, L., E. Pianta, and C. Girardi (2002, January). MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India.

Bentivogli, L., E. Pianta, and F. Pianesi (2000). Coping with lexical gaps when building aligned multilingual wordnets. In *Proceedings of LREC-2000*, Athens, Greece, pp. 993–997.

Corbett, G. G. (1991). *Gender*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Language, Speech and Communication. MIT Press.

Hellinger, M. (2001). English – gender in a global language. See Hellinger and Bußman (2001), pp. 105–114.

Hellinger, M. and H. Bußman (Eds.) (2001). *Gender across languages (volume I)*, Volume 9 of *Impact: Studies in language and society*. Amsterdam/Philadelphia: John Benjamins.

Janssen, M. (2004). Multilingual lexical databases, lexical gaps and SIMuLLDA. *International Journal of Lexicography 17*(2), 137–154.

Miller, G. A. (1990). Nouns in WordNet: a lexical inheritance system. *International Journal of Lexicography 3*(4), 245–264.

Tobin, Y. (2001). Gender switch in Modern Hebrew. See Hellinger and Bußman (2001), pp. 177–198.

Vossen, P. (2004). EuroWordNet: a multilingual database of autonomous and language-specific WordNets connected via an inter-lingual-index. *International Journal of Lexicography 17*(2), 161–173.