

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 March 2007 (15.03.2007)

PCT

(10) International Publication Number
WO 2007/029240 A2

(51) International Patent Classification:
G06F 19/00 (2006.01)

(21) International Application Number:
PCT/IL2006/001027

(22) International Filing Date:
5 September 2006 (05.09.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0518006.2 5 September 2005 (05.09.2005) GB

(71) Applicants and

(72) Inventors: **MELNICK, Einat, H.** [IL/IL]; 17 Binyamin Mitudela Street, 69548 Tel-aviv (IL). **MELNICK, Geofrey, L.** [GB/IL]; 17 Binyamin Mitudela Street, 69548 Tel-aviv (IL). **COHEN, David** [IL/IL]; 13 Rabina Street, Apt. # 29, 69395 Tel-aviv (IL).

(74) Agents: **G.E. EHRLICH (1995) LTD.** et al.; 11 Menachem Begin Street, 52 521 Ramat Gan (IL).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIGITAL UNIVERSAL LANGUAGE SYSTEM

(57) Abstract: A digital, translingual sense code and a system of operators are presented, for generating a digital, universal language system, which is substantially unequivocal. The digital language system may be used a translation-ready format, from which translation to any language may be automatic. Additionally, it may also be used for information retrieval and data acquisition, and as an aide in language acquisition. Furthermore, it may be used for tagging a natural language, for word sense disambiguation.

WO 2007/029240 A2

DIGITAL UNIVERSAL LANGUAGE SYSTEM

FIELD OF THE INVENTION

5 The present invention relates to a digital, translingual sense code and a system of operators, for generating a digital, universal language system, which is substantially unequivocal.

BACKGROUND OF THE INVENTION

10 For machine translation to be precise, a clear-cut correspondence is necessary between the source and target. Yet, words often have many senses, and the correspondence between the source and target is ambiguous and elusive - far from clear-cut.

15 One reason for the different senses is that language develops and grows by association.

20 For example, the word, verse - a line of poetry, comes from the Latin past participle of *vertere*, to turn, the association being that the lines of a poem resemble the lines a plough forms in a field, as it turns the soil. Many additional meanings to verse have developed, associated with a line of poetry, for example, verse as a poem, verse
25 as a metrical or rhymed composition distinct from prose, verse as poetry in general, verse as the work of a poet, and verse as a metrical writing without depth or artistic merit. Yet, the association between a line in a poem and a ploughed field, which has led to all these senses, is culture dependent, and could have developed only in an agricultural community; it would have made no sense to nomads. When nomads think
of a line, they think of a caravan – a single file of pack animals; indeed, in Hebrew, a line in a poem and a caravan are of the same root.

30 The word, bank, as a financial establishment, comes from Old Italian *banca*, a bench, a moneychanger's table, the association being that bank transactions are formed across a table. And while bank, originally referred to a place of storage of money, the association has now been expanded to include places of storage of data, blood, and other materials. Yet, the association between a bank and a bench is also culture dependent, and could have only developed where benches were used for transactions.

Associations are natural to the human mind, and language continues to grow and develop by new associations, constantly generating new senses to existing words. For example, to milk as to draw nourishing fluid from a teat or udder has led for example, to the following: to milk venom from a snake, to milk a witness for
5 information, or to milk money and benefits from someone. With the advent of nuclear medicine, a new sense developed, to milk Tc-99m from a technetium generator, for producing radiopharmaceuticals, such as Tc-99m-sestamibi, or Tc-99m-Teboroxime.

At times the new sense is momentary, used in analogy, and applying for a
10 particular task. At other times, the new sense catches on and becomes widespread. Yet, senses that grow from associations are both varied and fluid; it is practically impossible to catalog in a dictionary all senses associated with a word.

Another reason for the different senses is different origins.

For example, date, as a sweet, edible, oblong fruit, is from the Greek daktulos,
15 finger, the date fruit having a finger shape. Date, as a time defined by day, month, and year, is from the Latin data, issued (in Rome), on a certain day.

Similarly, bank, as a natural incline or the slope adjoining a river, is of Scandinavian origin, while bank, as a financial establishment, is of French and Old Italian origin.

20 In translation, one converts from a word associated with senses developed in one ancient system of origins and associations to a word associated with senses developed in another ancient system of origins and associations. The results may be unknown and unpredictable.

25 SUMMARY OF THE INVENTION

The present invention relates to a digital, translingual sense code and a system of operators, for generating a digital, universal language system, which is substantially unequivocal. The digital language system may be used a translation-ready format, from which translation to any language may be automatic.
30 Additionally, it may also be used for information retrieval and data acquisition, and as an aide in language acquisition. Furthermore, it may be used for tagging a natural language, for word sense disambiguation.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable
5 methods and materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Implementation of the method and system of the present invention involves performing or completing selected tasks or steps manually, automatically, or a
10 combination thereof. Moreover, according to actual instrumentation and equipment of preferred embodiments of the method and system of the present invention, several selected steps could be implemented by hardware or by software on any operating system of any firmware or a combination thereof. For example, as hardware, selected steps of the invention could be implemented as a chip or a circuit. As software,
15 selected steps of the invention could be implemented as a plurality of software instructions being executed by a computer using any suitable operating system. In any case, selected steps of the method and system of the invention could be described as being performed by a data processor, such as a computing platform for executing a plurality of instructions.

20

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 schematically illustrates a domain of senses and a domain of words, as
25 defined by the present invention;

Figure 2 schematically illustrates the construction of a translingual sense code and a translingual-sense-code lexicon, in accordance with preferred embodiments of the present invention;

Figure 3 illustrates a mapping of English dictionary entries, which may be
30 words, phrases, or idioms, into a translingual sense code, for extracting sense stems in accordance with preferred embodiments of the present invention;

Figure 4 is a transformation of Figure 3 to the domain of senses, in accordance with preferred embodiments of the present invention; and

Figure 5 illustrates a system for expressing the translingual sense code in digital format, using a vector of "n" natural numbers, in accordance with an embodiment of the present invention.

5 DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention relates to a digital, translingual sense code and a system of operators, for generating a digital, universal language system, which is substantially unequivocal. The digital language system may be used a translation-ready format, from which translation to any language may be automatic.
10 Additionally, it may also be used for information retrieval and data acquisition, and as an aide in language acquisition. Furthermore, it may be used for tagging a natural language, for word sense disambiguation.

The principles and operation of the universal language system according to the present invention may be better understood with reference to the drawings and
15 accompanying descriptions. But it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is
20 for the purpose of description and should not be regarded as limiting.

Words and Senses - The Fluid and the Discrete

Reference is now made to Figure 1, which schematically illustrates a domain of senses and a domain of words, as defined by the present invention. Senses may be
25 regarded as discrete entities, but words, the tools we use to relate to them, are fluid, with a spread of senses about each word, the senses spilling over and mingling with others. Humans have invented words to express senses, which seem to be somehow registered in their minds and reachable by multiple associations. In fact, it may be that the very spread in senses to a word provides the associations, that allows us to
30 reach a specific sense register we seek.

In consequence, the relationship between words and senses is such that one can generally find a proper word to describe a specific sense. But expressing words in terms of all their senses is a challenging and nearly impossible task.

Figure 1 illustrates this, for example, in reference to the word, "bank," which has a first "sense spread" about it, relating to holding reserves, and a second, relating to natural slopes. Although many discrete senses can be noted, for example, data bank, blood bank, piggy bank, and others, it is clear that many other senses may exist, or may be constructed and understood. Yet, delineating all of them is unlikely. There will probably be many senses that will be overlooked.

Figure 1 introduces a domain of words 10 and a domain of senses 20, and it is suggested here that they are very different in nature:

The domain of words 10 is a domain of fluid entities, and it may not be possible to describe a given word by all its senses.

The domain of senses 20, on the other hand, is a domain of discrete entities, and any given sense can be matched with words that describe it.

Thus, the success of the description depends on the vantage point: are we in the domain of words, looking across, at senses, or are we in the domain of senses, looking across, at words.

Construction of a Translingual-Sense-Code Lexicon

Defining a Sense Stem

The present invention introduce a new concept – a sense stem, analogous to a word stem. Yet, for the sake of clarity, a few definitions are offered:

a word, as referred to here, relates to speech sounds or a portion of text, which form a unit that communicates one or several meanings, the unit not being divisible into smaller units that communicate meanings. Each meaning is associated with a function, for example, a noun, a verb, a preposition, a conjunction. Additionally, each meaning may be associated with other attributes, such as case, gender, number, tense, person, mood, voice and others. For example, the word "sorter" may be a noun representing a human, or a noun representing an inanimate object.

a word stem, as referred to here, relates to the part of a word that remains substantially unchanged upon inflection. The stem has no function, as such, but is associated with the meanings of the inflected forms. For example, a stem "driv" is associated, among others, with *driver* and *driving*, in the following:

1. a *driver*, as a mechanical element for imparting motion to a second element, the first element *driving* the second piece into place;
2. a *driver*, as a person *driving* a motor vehicle; and
3. a *driver*, as an electronic circuit or software element that supplies input,
5 *driving* another electronic circuit.

a sense, as referred to here, relates to a single meaning of a word or phrase. A sense must be associated with a function, and possibly with other attributes. The sense stem, function and other attributes, together, define the single, specific meaning.

a sense stem, is a new concept, introduced by the present embodiments. It is
10 analogous to a word stem, and like the word stem, it has no function of its own but is associated with the functions of the inflected forms. It is different from a word stem in that it relates to a single sense. In other words, the sense stem for "driv," as associated with a *driver* - a person *driving* a motor vehicle is subtly but distinctly different from the sense stem for "driv," as associated with a *driver* - an electronic
15 circuit that supplies input, *driving* another electronic circuit.

Components of the Translingual-Sense-Code Lexicon

Reference is now made to Figure 2, which schematically illustrates the construction of a translingual sense code 49 and a translingual-sense-code lexicon 46,
20 in accordance with preferred embodiments of the present invention. Certain definitions are required:

The Sense stem 45: the sense stem 45 is a portion of the translingual sense code 49, which is independent of function and other attributes, and which remains unchanged upon inflection. On its own, the sense stem 45, analogous to the word stem
25 "driv" does not have a specific meaning.

In accordance with a preferred embodiment of the present invention, the sense stem 45 is a natural number. Examples of a sense stem may be 02300 or 327.

The Attribute 43: the attribute 43 or attributes 43 are the portions of the translingual sense code 49, which are expressed by inflection, and which include at
30 least a function, e.g., noun, verb, adjective, and the like, and preferably, additional features, such as case, gender, number, tense, person, mood, voice and the like. When a sense stem is inflected in accordance with its attributes, it has a specific meaning.

In accordance with a preferred embodiment of the present invention, the attributes 43 are described as natural numbers, and inflection is defined by the morphology rules, below.

The Morphology Rules 47: The morphology rules 47 define how a sense stem is to be inflected so as to express various attributes. It is a system for combining the sense stem 45 and the attributes 43, to form the translingual sense code 49.

In accordance with a preferred embodiment of the present invention, the sense stem 45 and each of the attributes 43 are described as natural numbers, while the morphology rules 47 define the order of these in a vector of n natural numbers, A(1),A(2)...A(n). The morphology rule may define the sense stem as the first position, A(1), the function as the second position, A(2), and so on.

The Translingual Sense Code 49: The translingual sense code 49 is an unequivocal word equivalent, formed of a sense stem 45, inflected to express specific attributes 43, in accordance with the morphology rules 47.

In accordance with a preferred embodiment of the present invention, the translingual sense code 49 is a vector of n natural numbers, for example, A(1),A(2)...A(n). A specific example of a translingual sense code 49 may be 04110,1,3,1,0,0,0,1, which in Example 2 below, means, a human sorter - a person who sorts things.

Syntax Rules 41: The syntax rules 41 specify how the unequivocal word equivalents of the translingual sense codes 49 are joined to form phrases, clauses and sentences.

In accordance with one embodiment of the present invention, the syntax rules 41 may be a predetermined order of arranging the vectors of natural numbers, using common punctuation marks, to form phrases, clauses and sentences.

In accordance with a preferred embodiment of the present invention, the syntax rules 41 are numerical operators, for example, +, X, :, *, (), and others, which define relations between the vectors of natural numbers, to form phrases, clauses and sentences.

The Translingual-Sense-Code Lexicon 46: The translingual-sense-code lexicon 46 is an unequivocal, written language, also referred to as a universal language system, in which sense stems 45, are combined with various attributes 43, through the morphology rules 47, to form unequivocal words or translingual sense

codes 49, and these are combined via syntax rules 41, to form the form phrases, clauses and sentences.

The translingual-sense-code lexicon 46 may be used as a translation-ready format, from which translation to any language may be automatic. Additionally, it
5 may also be used for information retrieval and data acquisition, and as an aide in language acquisition. It may further be used for tagging, for word sense disambiguation.

The translingual-sense-code lexicon 46 as described herein is generative, as it can create unequivocal word equivalents, by combining sense stems and attributes,
10 which may not exist in any language. As such, it may be of relevance in discussions of The Generative Lexicon, Pustejovsky (1995), Bouillon & Busa, (eds.) (2001).

Extracting the Sense Stem

There remains the question of how the sense stem 45 is to be extracted.

15 Reference is now made to Figure 3, which illustrates a mapping 40 of English dictionary entries 42, which may be words, phrases, or idioms, into the translingual sense code 49, for extracting sense stems 45, in accordance with preferred embodiments of the present invention. The dictionary entries 42 may be grouped into synonym clusters 44, and (or) provided with definitions.

20 Additionally, each dictionary entry 42 is noted for its attribute 43, such as function, person, number and the like. The senses are described as the translingual sense code 49, preferably of natural numbers, for example, where A(1) denotes the sense stem 45, and A(2) – A(n) denote the other attribute 43. In the example of Figure 3, the dictionary entries are transitive verbs in the infinitive forms, so A(2) = 2
25 (See Figure 5, hereinbelow), and the sense is described as the translingual sense code 49 of A(1), A(2).

The translingual sense code 49, describing the senses, is preferably linked to synonym clusters 48 and (or) definitions, in other languages, to form the translingual system. For example, the word “order” as a transitive verb, in the infinitive form,
30 may be clustered with “arrange” and “organize,” assigned the natural number 04100 as the sense stem 45, so that the translingual sense code 49 is 04100,2 and linked to synonym clusters 48 in Hebrew and German, of substantially the same sense. Naturally, other languages may be included as well.

It will be appreciated that monolingual definitions may be used, in place of, or in additional to the synonym clusters, for example, "to order - to put into a methodical arrangement." The monolingual definitions may be based on various relationships, for example, goose – a large waterfowl, intermediate between swans and ducks, or a gander – an adult male goose, and other relationships, as known.

Phrases and idioms may be grouped into synonym clusters, in the same manner as words, for example, the idiom "sleep on it" may be defined by another idiom, "think it over" and (or) by the words "consider," and "reflect."

Figure 3 is a representation of the domain of words 10 of Figure 1, hereinabove.

Reference is now made to Figure 4, which is a transformation of Figure 3, to the domain of senses 20 of Figure 1, in accordance with preferred embodiments of the present invention.

Figure 4, illustrates a transform from the domain of words 10 to the domain of senses 20, in accordance with the present invention. Figure 4 is arranged by the sense stems 45 and the translingual sense code 49, describing the senses, linked to their associated synonym clusters and (or) definitions 44 and 48 of all of various languages, for example, English, Hebrew, and German.

Figure 4 represents a translingual sense dictionary 50, and brings the senses out into the open, coded in a language-independent, manner, preferably, digitally, so as to replace the multiple-association register of the mind, illustrated in Figure 1 with the unequivocal, translingual sense code 49.

As a matter of definition, each single sense of the translingual sense code 49 may also be referred to as a translingual-sense-code entry.

25

EXAMPLES

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

Reference is now made to the following examples, which together with the above descriptions, illustrate the invention in a non limiting fashion.

Example 1: The Translingual-Sense-Code 49 as a Vectors of Natural Numbers

Referring is now made to Figure 5, which illustrates a system for expressing the translingual sense code 49 in digital format, using vectors of “n” natural numbers, A(1), A(2)..... A(n), in accordance with an embodiment of the present invention.
 5 A(1), A(2)..... A(n), in accordance with an embodiment of the present invention.
 Each number “n” is expansible, as necessary, being of a single digit, double digits, and so on.

According to the present example:

- the first position A(1) defines the sense stem 45;
- 10 • the second position A(2) defines the function, so $A(2) = 1$, inflects the sense stem to mean a noun of the specific sense, and so on;
- the third position A(3) defines the person, so $A(3) = 1$, inflects the sense stem to express first person. Note, that in some languages, this inflection applies only for pronouns, in others, it applies to verbs; and still in others, to verbs, adjectives, and
 15 even prepositions. The person inflection in the translingual sense code 49 is made available, regardless of function, for languages that require it, and it may be ignored by languages that do not require it.
- the fourth position A(4) distinguishes between male, female, and neuter, in the singular and plural forms. Again, it is made available, regardless of function, for
 20 languages that require it, and it may be ignored by languages that do not require it.
- the fifth position A(5) is used to indicate that a phonetic translation is required, for example, for a name. In some cases, the entry of the fifth position is of the phonetic translation, in accordance with phonetic symbols, for example as described in comment 2, below. Note, in this instance, there will be no sense stem. The town
 25 name Chelmsford, may be expressed, for example, as follows: $A(1) = 0$, $A(2) = 1$, $A(3) = 3$, $A(4) = 3$, $A(5) = \text{ch}^{\text{e}}\text{l}^{\text{m}}\text{z}^{\text{f}}\text{r}^{\text{d}}$, or, 0,1,3,3, $\text{ch}^{\text{e}}\text{l}^{\text{m}}\text{z}^{\text{f}}\text{r}^{\text{d}}$. It will be appreciated that a universal system for pronunciations, as suggested in comment 2 hereinbelow, is preferred.
- the sixth position A(6) indicates the word may be colloquial, legal, or
 30 otherwise unusual.
- the seventh position, A(7), expresses the case, when the function A(2) is a noun, indicates if the predicate, when the function A(2) is a verb, and so on.

Other values are similarly apparent from Figure 5.

The following comments relate to the superscripts of Figure 5:

1. Noun phrases and verb phrases may be linked, using A(11) or A(12).
2. A digital phonetic code, which includes all the sounds and vowels of all the languages is required, for phonetic translations, such as of names. New sounds need to be introduced to each language to cover these, in a systematic manner. For example, kh may be defined in English for the Spanish "J". When A(5) includes the digital phonetic code, the word will be translated phonetically, using the target-language alphabet, with a full set of sounds.
3. Noun types may relate to: 1=human, 2=animal, 3=plant, 4=object, 5=abstract, 6=action, 7=place, 8=time, and others.
4. A basic tense structure of: 0=infinitive, 1=past, 2=present; 3=future, and 4=imperative, may be used. Alternatively, an expansible system, which describes the particular tenses of a specific language, may be used, for example, noting the language as the first digit, and the tenses with the other digits. For example, let English be denoted by the first digit "1", and the English tenses described as 11=Eng, past simple, 111=Eng, past perfect, 1101=Eng, past continuous, 1111= Eng, past perfect continuous, 12= Eng, present simple, 121= Eng, present perfect, 1201= Eng, present continuous, 1211= Eng, present perfect continuous, and so on.
5. An expansible system may be used to denote affirmative, negative, interrogative, and negative interrogative (e.g., did you not...).
6. An active adjective may relate to washing, as in a washing machine, a passive adjective may relate to washed, as in washed clothes, a reflexive adjective may relate to sleeping, as in a sleeping man, an active-able adjective may relate to one or that, which is cable of washing, a passive-able adjective may relate to one or that, which can be washed, and a reflexive-able adjective may relate to one who is cable of self-washing (i.e., self cleaning). Naturally, other forms may also be defined.
7. Where the translingual sense code 49 represents a sense with attributes that do not exist, within a single word or term, in certain languages, definitions or phrases, based on the attributes may be produced by the operating utility, in those languages.
8. The linking indices, of A(11) – A(12) serve to overcome different syntax order in different languages. They form chains, which retain their meanings even as the order of words change from one language to another. For example, in "I wish to go

by-car, to-the-cinema," and "I wish by-car, to-the-cinema go," the prepositional phrases are linked, and will remain so in the different languages, in spite of the different order. Similarly, the order of noun-adjective, or adjective-noun, is unimportant, as the nouns and adjectives are marked by their functions and linked by

5 A(11)

9. The translingual sense code need be described only to the last non-zero term, e.g., if the last non-zero term is A(8) it need have only 8 natural numbers.

In a way, the translingual sense code 49 operates as a checklist for verifying that all information relevant to the decision making of the operating utility will be available.

10 In general, the translingual sense code 49 may be formed automatically, by the operating utility. But where information is lacking, human input is sought.

It will be appreciated that another coding system, based on the Latin alphabet, the Greek alphabet, Roman numerals, real numbers, or another system as known, may be used. For example, with Latin letters, a sense stem 04110 could be represented as

15 TjKn and a complete translingual sense code may be expressed, for example, as "TjKn,I,l,ee,w." The advantage of natural numbers is that they are thrifty in terms of digital storage space, since they do not require conversion to ASCII.

Additionally, the vector of natural numbers may include any number of natural numbers, so n may be 9 or 15. While it is advantageous for n to be 8, 16, or any

20 other number, which can be expressed as 2 to some factor, that is not necessary.

It will be appreciated that many other tables may be created, for systematically defining attributes and possibly also linkages between terms.

Example 2:

25 The manner of generating the translingual sense code 49, for example, using Figure 5, is illustrated in reference to the paragraph below:

"Sam was a sorter. He sorted files for a living. But the sorted files got to him. He was tired of sorting. So he bought a sorting machine to sort his files. And he promised himself that he would never sort files again."

30 Letting a natural number 04110 represent the sense stem 45, as relating to sorting - arranging according to characteristics, the following constructions can be made:

1. *sorter* in " Sam was a *sorter*"

In accordance with the preferred embodiment of the present invention, the translingual sense code 49 that forms the unequivocal word equivalent for a human sorter, as in "Sam was a sorter," may be constructed in accordance with the definitions of Figure 5, as follows:

- A(1) denotes the sense stem, for example, 04110;
 - A(2) denotes a noun;
 - A(3) denotes 3rd person;
 - A(4) denotes male;
 - A(8) denotes human;
- sorter* = 04110,1,3,1,0,0,0,1.

2. *sorted*, in "He *sorted* files for a living"

- A(1) denotes the sense stem, 04110;
- A(2) denotes a transitive verb;
- A(3) denotes 3rd person;
- A(4) denotes male;
- A(7) denotes predicate;
- A(8) denotes past tense.
- A(9) denotes: active.

sorted = 04110,2,3,1,0,0,1,1,1.

3. *sorted* in "But the *sorted* files got to him"

- A(1) denotes the sense stem, 04110;
- A(2) denotes adjective;
- A(3) denotes 3rd person;
- A(4) denotes male;
- A(7) denotes first adjective;
- A(9) denotes adjective in a passive form.

sorted = 04110,3,3,1,0,0,1,0,2.

Although "sorted," as an adjective, does not appear in most English dictionaries, the sense morphology 47 enables one to generate a meaning for "sorted" as an adjective, by combining the in-context sense stem of A(1) with the in-context function of A(2), making it possible to create senses as necessary and as relevant. Similarly, 5 attributive nouns, for example, "city," in city lights, or "health," in health care, will be assigned an adjective function, which is their in-context function. With the sense morphology 47, all nouns can be verbed, and anything can be described adverbially.

The sense morphology 47 makes it possible to clearly distinguish between senses of different attributes, for example:

- | | | |
|----|--|------------------------|
| 10 | • noun: a human sorter: | 04110,1,3,1,0,0,0,1; |
| | • noun: a machine sorter: | 04110,1,3,1,0,0,0,4; |
| | • noun: an action, sorting: | 04110,1,3,1,0,0,0,6; |
| | • adjective: sorted (e.g. sorted files): | 04110,3,3,1,0,0,1,0,2; |
| | • adjective: sorting (e.g. a sorting machine): | 04110,3,3,1,0,0,1,0,1; |
| 15 | • adjective: capable of being sorted: | 04110,3,3,1,0,0,1,0,5; |
| | • adjective: capable of sorting: | 04110,3,3,1,0,0,1,0,4; |
| | • a transitive verb: to sort: | 04110,2. |

As another example, letting a natural number 05230 be the sense stem 45, as relating to washing - cleansing by wetting thoroughly with water, to carry off foreign 20 matter, the following senses can be constructed:

- | | | |
|----|---|------------------------|
| | • adjective: capable of being washed (washable): | 05230,3,3,1,0,0,1,0,4; |
| | • adjective: capable of washing (washingable): | 05230,3,3,1,0,0,1,0,3; |
| | • adjective: capable of self-wash (selfwashable): | 05230,3,3,1,0,0,1,0,5. |
| | • a transitive verb, to wash (e.g., clothes): | 05230,2. |
| 25 | • an intransitive verb, to wash (oneself): | 05230,8. |

As it happens, some of these senses do not have specific words or terms in some languages. For example, in English, there is no specific word or term for "capable of washing," or for "capable of self-wash," But the translingual sense code 49 provides 30 these meanings nonetheless, and it is in this sense that it is generative. Upon

translation to natural languages, where no equivalent term is available, a definition may be used.

Example 3 – Syntax Rules, in Accordance with a First Embodiment

5 Given that the translingual sense code 49 are unequivocal word equivalents, they may be combined to form phrases and sentences, using any known syntax rules, for example, those of the English languages.

 Accepting the order: subject, predicate, direct object, as syntax rules 41, the sentence, “He sorted files,” can be expressed by the translingual-sense-code lexicon
10 46, and specifically, by the digital lexicon 46, as follows:

“He” “He” needs no sense stem, being defined by the other attributes.

“sorted” The transitive verb “sorted,” in the past tense, has been described in the second example of digital sense morphology, above.

“files” Let the natural number for the sense stem 45 associated with “file - a
15 collection of papers arranged in a folder,” be 06750.

Thus, “He sorted files” =

= 0,1,3,1,0,0,1,1,2 04110,2,3,1,0,0,1,1,1 06750,1,3,30,0,0,0,4.

Example 4 – Syntax Rules, in Accordance with a Preferred Embodiment

20 In accordance with the preferred embodiment, numerical operators are used to define syntax rules, for example, as follows:

- A relationship between subject and predicate may be expressed by X,

25 Example: “he works” may be expressed, for example, as “he X works,” or as “works X he.”

Wherein: the subject and predicate remain juxtaposed, but their order is unimportant.

- A relationship specifying equivalence may be expressed by =,

30 Example: “they are friends” may be expressed, for example, as “they X are = friends,” or as “friends = they X are.” or as “friends = are X they.”

Wherein: the subject and predicate remain juxtaposed, but their order is unimportant.

- A relationship between subject-predicate combination and an object may be expressed by :,

Example: “he took it” may be expressed, for example, as “he X took : it,” or as “took X he : it,” or as “it : he X took,” and even as “it : took X he.”

5 Wherein: the subject and predicate remain juxtaposed, but their order is unimportant.

- A relationship between an object and a verb may also be expressed by X,

Example: “he saw me come” may be expressed, for example, as “he X saw : me X come,” or as “me X come : he X saw,” or as “come X me : he X saw,” or as “come X me : saw X he,” or by other similar combinations.

10

Wherein: the subject and predicate remain juxtaposed, and the object and its associated verb remain juxtaposed, but the order is unimportant.

- A relationship between an adjective and a noun may be expressed by *,

15 Example: “good book” may be expressed, for example, as “good*book,” or as “book*good.”

Example: “it is late” may be expressed, for example, as “it X is * late,” or as “late * it X is,” or as “late * is X it.”

- A relationship between an adverb and a verb may be expressed by #,

Example: “She sang beautifully” may be expressed, for example, as “she X sang # beautifully,” or as “beautifully # she X sang,” or as “beautifully # sang X she,” or as or as “sang X she # beautifully.”

Wherein: the subject and predicate remain juxtaposed, but their order is unimportant.

25 Thus, the system of the present example does not allow “sang # beautifully X she,” requiring that the subject-predicate combination take precedence over other combinations. It will be appreciated that other syntax systems may be devised, with different restrictions, provided they are consistent, throughout.

- A relationship between an adverb and an adjective may be expressed by #,

30 Example: “highly useful remark” may be expressed, for example, as “highly # useful * remark,” or as “useful # highly * remark,”

Wherein: in accordance with the present example, the adverb-adjective combination takes precedence over the adjective-noun combination. It will be appreciated that a syntax system with an opposite rule is similarly possible, provided the rules are consistent, throughout.

5

- A relationship between an adverb and a clause may also be expressed by #, the clause being in parenthesis,

Example: “however, it is late” may be expressed, for example, as “however # (it X is * late),” or as “(it X is * late) # however.”

10 Wherein: in accordance with the present example, the adverb modifying a whole clause may not be inserted in the middle of the clause, so that “it X is # however * late,” is not acceptable in accordance with the present example, as it creates a confusion regarding what specifically “late” relates to. It will be appreciated that other syntax systems with different rules may be applied.

15

- A relationship between a group of adjectives may be expressed by +, the group being in parenthesis,

Example: “good, enjoyable book” may be expressed, for example, as “(good + enjoyable) * book,” or as “(enjoyable + good) * book,” or as “book * (enjoyable + good),” or as “book * (good + enjoyable).”

20

Wherein: in accordance with the present example, the group of adjectives are not broken up, so that “enjoyable * book * good” is not accepted under the present system. It will be appreciated that other syntax systems with different rules may be applied.

25

- A relationship between a group of adverbs may be expressed by +, the group being in parenthesis,

Example: “she sang beautifully, melodically” may be expressed, for example, as “she X sang # (beautifully + melodically),” or as “(beautifully + melodically) # sang X she,” or as a similar combination.

30

Wherein: in accordance with the present example, the subject and predicate remain juxtaposed and so is the group of adverbs. It will be appreciated that other syntax systems with different rules may be applied.

- A relationship between a group of nouns may be expressed by +, the group being in parenthesis,

Example: “Dan and Jim left” may be expressed, for example, as “(Dan + Jim) X left,”
5 or as “left X (Jim + Dan),” or as a similar combination.

Wherein: in accordance with the present example, the group of nouns remain together, in parenthesis. It will be appreciated that other syntax systems with different rules may be applied.

- 10 • A relationship between a group of verbs may be expressed by +, the group being in parenthesis,

Example: “they ate and slept” may be expressed, for example, as “they X (ate + slept),” or as “(slept and ate) X they.”

- 15 Wherein: in accordance with the present example, the group of verbs remain together, in parenthesis. It will be appreciated that other syntax systems with different rules may be applied.

- A relationship between a conjunction and a clause may be expressed, for example, as ^, and the clause may be in parenthesis.

- 20 Example: “while they slept soundly...” may be expressed as “while ^ (they X slept # soundly)...,” or as “while ^ (soundly # they X slept),” or as “(soundly # they X slept) ^ while...,” or to similar combinations.

- 25 Wherein: in accordance with the present example, “while” relates to the time they slept soundly, regardless of its position. It will be appreciated that other syntax systems with different rules may be applied.

- A relationship between a main clause and a subordinate close may be expressed, for example, as /,

- 30 Example: “while they slept soundly, she cleaned” may be expressed, for example, as “while ^ (they X slept # soundly) / she X cleaned,” or as “(they X slept # soundly) ^ while / she X cleaned,” or as “she X cleaned / while ^ (they X slept # soundly)” or as “she X cleaned / (they X slept # soundly) ^ while,” or as similar combinations.

Wherein: in accordance with the present example, “while” relates to the time they slept soundly, and the clause “while they slept soundly” remains the subordinate clause, regardless of the order of the two clauses. It will be appreciated that other syntax systems with different rules may be applied.

5

- A relationship between the components of noun phrases and noun clauses, as well as verb phrases may be maintained by square brackets, [],

Example: “that it is true was shown experimentally” may be expressed, for example, as “[that ^ (it X is = true)] X [was shown] # experimentally,” or as “experimentally # [that ^ (it X is = true)] X [was shown], or as “experimentally # [was shown] X [that ^ (it X is = true)],” or as “[was shown] X [that ^ (it X is = true)] # experimentally.”

10

Wherein: the subject and predicate remain juxtaposed, even when each is a phrase or a clause, but their order is unimportant.

15

It will be appreciated that many other operators and combinations of operators are similarly possible.

Example 5 – Syntax Operators with the Translingual Sense Code 49

In accordance with the preferred embodiment, using operators for expressing syntactic relations, the sentence of Example 3, “He sorted files,” may be expressed as:

20

0,1,3,1,0,0,1,1,2 X 04110,2,3,1,0,0,1,1,1 : 06750,1,3,30,0,0,0,4.

Example 6 – Translation Ready Format

Embodiments of the present invention may be employed for providing translation ready formats, which are substantially unequivocal, for automatic translation to a plurality of languages. For example, a computer system may be employed, containing a set of instructions, for

receiving a natural source language, comprising words and syntax rules; and

parsing the natural language, based on the syntax rules; and

converting the words to senses, based on a data base comprising:

senses, coded in a universal, language-free format; and

definitions, associated with each sense;

re-writing the natural source language as a universal language system.

Additionally, interactive human input may be employed for converting the words to senses, and possibly also for parsing.

Furthermore, the computer system may be employed for translating the universal language system to any natural language, automatically.

As such, the universal language system, which is preferably digital, as described, becomes a translation-ready format for storing information, easily converted to any natural language, on demand.

* * *

It is expected that during the life of this patent many relevant universal language systems will be developed and the scope of the term universal language system is intended to include all such new technologies a priori.

As used herein the term “substantially” refers to $\pm 10\%$.

As used herein the term “about” refers to $\pm 30\%$.

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein

by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

WHAT IS CLAIMED IS:

1. A data base comprising:
 - senses, coded in a universal, language-free format; and
 - definitions, associated with each sense.
2. The data base of claim 1, wherein the senses are coded digitally.
3. The data base of claim 1 or 2, wherein the senses are coded digitally, as vectors of n natural numbers.
4. The data base of any one of claims 1 - 3, wherein each sense is coded at least as a distinct sense-stem portion and a distinct portion specifying a function.
5. The data base of claim 4, and further including distinct sense portions specifying additional attributes.
6. The data base of any one of claims 1 - 5, wherein the senses may be used as unequivocal word equivalents, in a universal language system.
7. The data base of any one of claims 1 - 6, wherein the senses may be used for tagging for word sense disambiguation.
8. The data base of any one of claims 1 - 7, wherein the senses may be used for information retrieval.
9. A method of written expression, using a universal language system, comprising:
 - providing senses, coded in a universal, language-free format; and
 - associating the senses by syntax rules, which define relationships between the senses,

so that the universal language system may be used for written expression, such as phrases, sentences, whole documents, and literary works, in a substantially universal and unequivocal manner.

10. The method of claim 9, wherein the syntax rules are operators.
11. The method of claim 9, wherein the universal language employs the data base, in accordance with any one of claims 1-8.
12. The method of claim 9, wherein the universal language is a digital universal language system.
13. A computer system, containing a set of instructions, for receiving a natural source language, comprising words and syntax rules; and parsing the natural language, based on the syntax rules; and converting the words to senses, based on a data base comprising:
 - senses, coded in a universal, language-free format; and
 - definitions, associated with each sense;re-writing the natural source language as a universal language system of any one of claims 9-12.
14. The computer system of claim 13 and further including interactive human input.
15. The computer system of any one of claims 13 – 14, and further including translating the universal language system to a natural language, automatically.
16. A computer-readable medium, containing a set of instructions, for receiving a natural source language, comprising words and syntax rules; and parsing the natural language, based on the syntax rules; and converting the words to senses, based on a data base comprising:
 - senses, coded in a universal, language-free format; and
 - definitions, associated with each sense;

re-writing the natural source language as a universal language system of any one of claims 9-12.

17. The computer-readable medium of claim 16 and further including interactive human input.

18. The computer-readable medium of any one of claims 16 – 17, and further including translating the universal language system to a natural language, automatically.

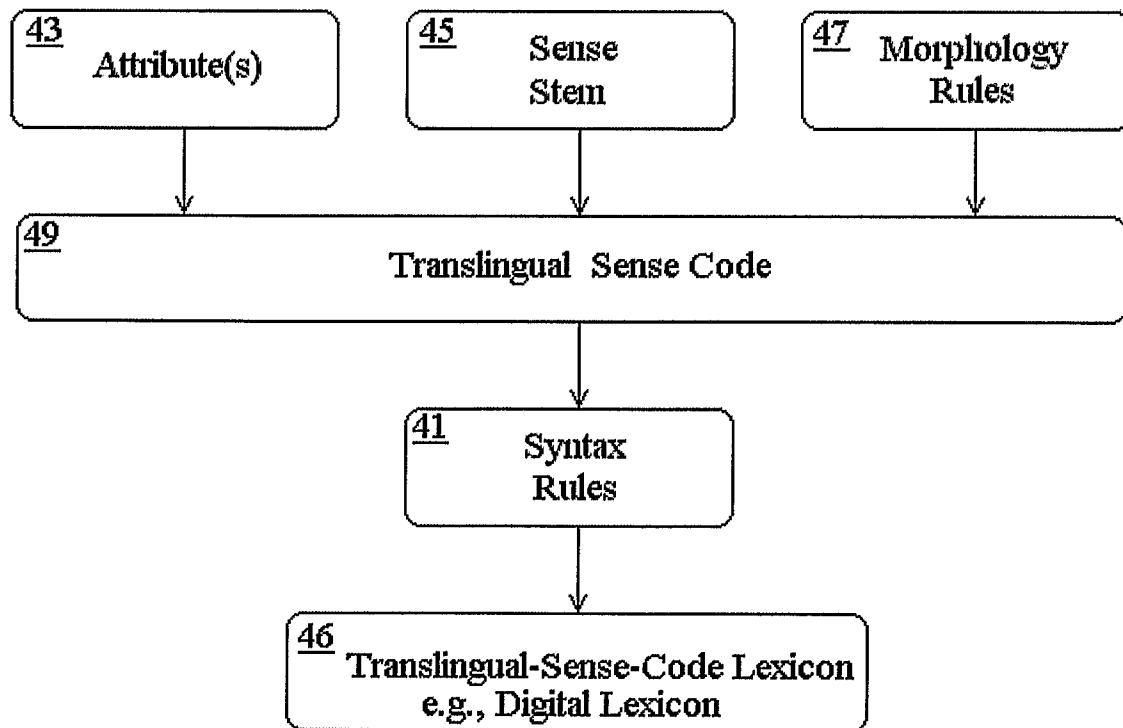
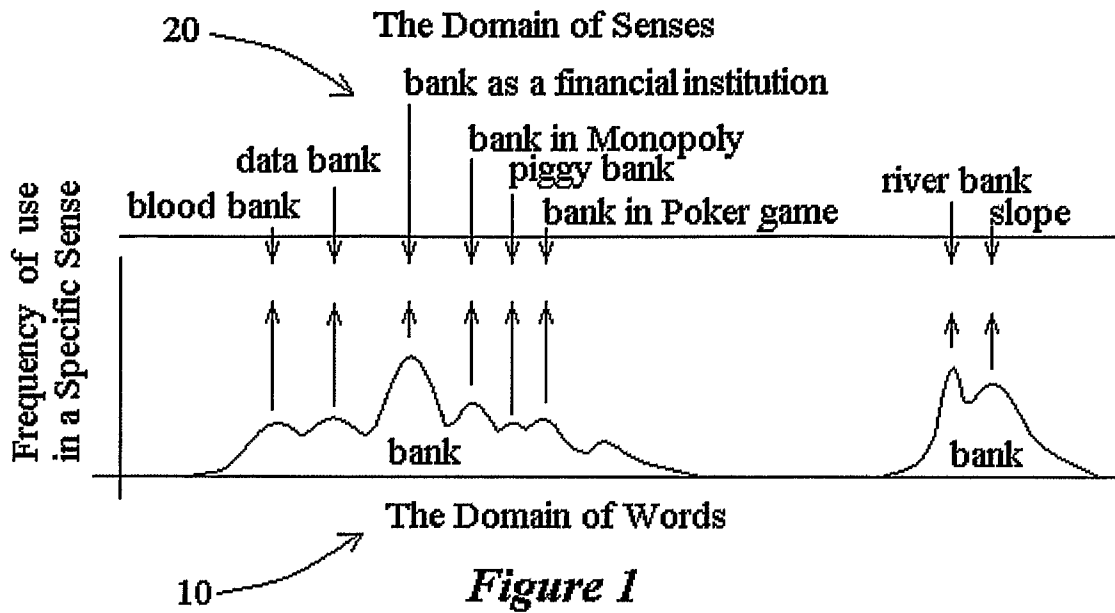


Figure 2

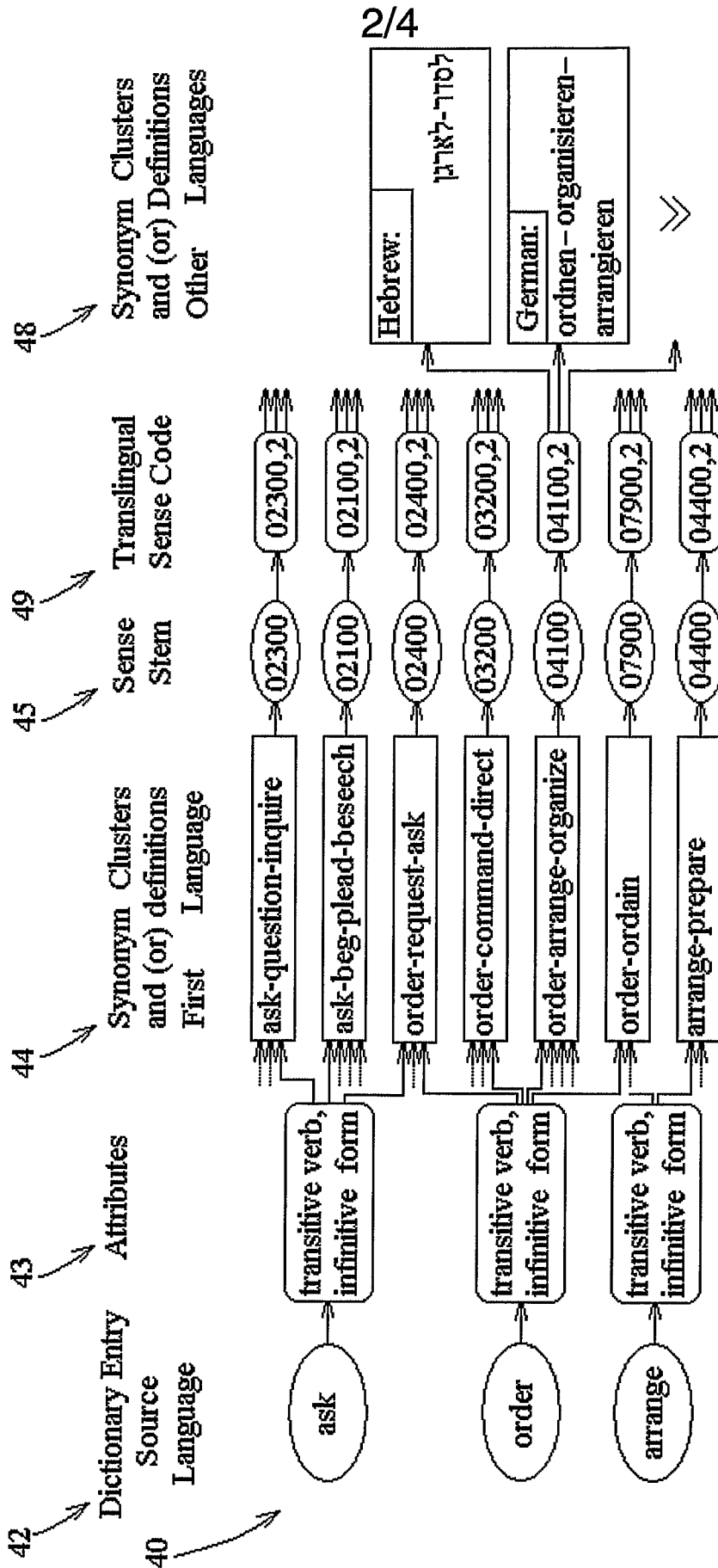


Figure 3

45 Sense Stem	49 Translingual Sense Code	44, 48 Synonym Clusters and (or) Definitions
04100	04100,2	English: order-organize-arrange Hebrew: לארגן - לדב German: ordnen-organisieren-arrangieren
04110	04110,2	English: order-organize-arrange - sort Hebrew: למיין - לארגן - לדב German: ordnen-organisieren-arrangieren - sortieren
04120	04120,2	English: order-organize-arrange- tidy up Hebrew: לתקן - לתקן - צורף - לארגן - לדב German: ordnen-organisieren-arrangieren-aufraeumen

50 →

Figure 4

Figure 5

natural number A(n)	Attribute 43						
A(1)	sense stem (e.g., 04110)						
A(2)	function						
	noun=1 ⁽¹⁾	tr. v.=2 ⁽¹⁾ int. v.=8	adj.=3	adv.=4	prep.=5 ⁽¹⁾	Conj. =6	article =7
A(3)	1=1 st person; 2=2 nd person; 3=3 rd person						
A(4)	1=male, 10=males, 2=female, 20=females, 3=neuter singular, 30=neuter plural						
A(5)	translate phonetically ⁽²⁾						
A(6)	1= archaic, 2=colloquial, 3=legal, 4-unusual, 5-idiomatic expression...						
A(7)	1= Subj. 2=obj.	1= predicate	1=1 st adj. 2=2 nd adj.	1=1 st adv. 2=2 nd adv.	1=1 st 2=2 nd prep. phrase	1=1 st 2=2 nd Subordi -nate phrase	1=a 2=the 3=said
A(8)	type ⁽³⁾	tense ⁽⁴⁾	possessive				
A(9)	1= name, 2= pro-noun.	voice e.g. 1=active 2= passive 3= reflexive 4= auxiliary ⁽⁵⁾	1= active, 2= passive, 3= reflexive, 4= active- able, 5= passive- able, 6= reflexive- able, 7= ful, 8=less ⁽⁶⁾⁽⁷⁾				
A(10)	relating to a patent element						
A(11) ⁽⁸⁾	links senses of a noun phrase by an index, if desired ⁽¹⁾						
A(12)	links senses of a verbal phrase by an index, if desired ⁽¹⁾						