Marcel Thelen, Maastricht State College of Higher Education,
School of Translation and Interpreting

Lexical Systems and Lexical Domains as Measures of Accessibility, Consistency and Efficiency of Lexical Information in Dictionaries[1]

ABSTRACT: This paper gives an account of a small-scale empirical study
on the accessibility, consistency and efficiency of lexical information in a
number of monolingual dictionaries. The measures of analysis and com-
parison are the Lexical System and the Lexical Domain, which are among
the most important cognitive/ semantic structuring principles of the the-
oretical lexicon in a grammar called "TCM" for "Two-Cycle Model of
Grammar". These principles may be of great help not only in the analysis
of dictionaries, but also for lexicographers and the users of dictionaries,
especially student translators.

1. General

Quite often a comparison of dictionaries seems to entail nothing more than just putting together actual/concrete formulations of definitions of a number of entries from the various dictionaries of investigation and looking for similarities and differences of form without going any deeper and trying to find underlying similarities and differences on a more abstract level. Such a method may be useful but is defective in a number of ways. Even though in such a method dictionary definitions are apparently regarded as formu- lations of meaning and thus as entities underlying actual lexical items, they are neverthe- less forms consisting of a sort of shorthand or condensed sentences with actual words functioning as their constituting elements. Dictionary definitions are, in fact, surface structure elements, and not deep structure entities. Comparing only forms may leave quite a few similarities and differences undetected, or even lead to wrong conclusions. What is lacking in such a method is an abstract framework for comparison. What is needed is a method of abstraction from actual forms to underlying concepts. This is not to say that one should look for so-called primitives of meaning in an effort to avoid the well-known and seemingly highly dangerous phenomenon of circularity between form and meaning. If the proper principles are applied, one need not be afraid of this.

In this paper, I will give an account of a comparison of a number of monolingual dictionaries on the point of lexical information for a number of lexical items. In this

comparison I applied a number of principles from the framework provided by a linguistic model called "TCM" (for "Two-Cycle Model of Grammar"), i.e. the Lexical System and the Lexical Domain. I will try to show that with this framework at least some parts of dictionaries may be assessed and, if necessary, improved on the point of accessibility, consistency and efficiency.

2. The notions of accessibility, consistency and efficiency

In this study, I shall treat the notions of accessibility, consistency and efficiency from the perspective of the users of dictionaries, and notably from the perspective of how informative individual instances of lexical information are for dictionary users.

By ACCESSIBILITY of lexical information I mean its "transparency" for the user: the more transparent or self-explaining the information is, the more accessible it will be for the user, i.e. the better will he understand and appreciate what is said. The fewer synonyms are given in a dictionary's lexical definitions, and the fewer look-ups are needed to trace their meaning(s), the greater will be its accessibility. If synonyms are given, and the meaning of these synonyms is incompatible with the rest of the lexical information and/or the examples given, accessibility will be zero. It goes without saying that here also the number of look-ups needed to find the lexical item required plays a role. For nouns, the notion of accessibility simply entails whether their lexical information can be "transformed back" to the LS of investigation, in other words, whether they belong to the LS of investigation.

By CONSISTENCY I mean that, for lexical items belonging to a particular group of lexical items that are one way or the other related in meaning, as much as possible one and the same for the user recognizable, and therefore consistent, framework of formulation of lexical information is given.

By EFFICIENCY I mean that a minimum length and content of actual lexical information are given, but long and explicit enough so as to achieve a maximum of accessibility and consistency and so as not to be too much of a strain for the user to read and comprehend: the most efficient lexical information is that information that uses as few definitional words as possible and as many as is necessary, and that is yet accessible and consistent.

3. Dictionaries used in the investigation

The main criterion for selection of the dictionaries of comparison was the actual learning practice of students of translation in the Department of English at the School of Translation and Interpreting of the Maastricht State College of Higher Education. Since students of translation are supposed to be(come) the most critical and scrutinous users of dictionaries and are helped best with lexical information that is as exhaustive and explicit as possible, it only seemed natural to concentrate on this group of users in my study and to take those dictionaries as the dictionaries of comparison that they use most frequently in the process of translation. For this study I concentrated on the decoding stage of the translation process from English to Dutch, i.e. on monolingual English dictionaries.

These dictionaries are: (1) COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY (Cobuild), (2) COLLINS ENGLISH DICTIONARY (Collins), and (2) OXFORD ADVANCED LEARNER'S DICTION-ARY OF CURRENT ENGLISH (Oaldce).

4. TCM: a Two-Cycle Model of Grammar

Originally developed by Alinei (1980), the most recent refinements, revisions and appli-cations of TCM were given in Thelen (1987a, 1987b, 1987c, 1990, 1991, and 1992 (in print)), and Thelen and Starren (1991). Its main principle is that the meaning of a lexical item is a prototypical concept and that this concept has a particular structure that resembles the syntactic structure of an actual sentence. Lexical items are condensed forms of under-lying conceptual-syntactic structures. Such structures underlying lexical items are nearly identical with the purely syntactic structures at the level of sentences. In such structures there are not only conceptual-syntactic categories such as SB (= Subject), PD (= Predi-cate), OB (= Direct Object), IO (= Indirect Object), LOC (= Locative), MANN (= Manner), etc., but also conceptual-semantic features or components, such as <human>, <eat>, <food> etc. Roughly speaking, SB is the category for the entity performing or undergoing the action or the entity being in the state identified with a verbal lexical item, PD the category for the action or state itself, etc. Conceptual-syntactic categories underlying lexical items have their parallel in the syntactic categories of sentences, and may be compared with Cases. Thus, SB is identical with Subject NP, etc. The conceptual-semantic components figuring in the underlying conceptual-syntactic structures are actual words in sentences. As a result the Grammar is doubled into two Cycles: the Lexical Cycle and the Sentence Cycle. Both Cycles have a deep structure and a surface structure, whereas the Lexical Cycle is "deeper" than the Sentence Cycle. The conceptual-syntactic struc-tures underlying lexical items are the output of the Lexical Cycle and are called "Inter-nalised Sentences or Phrases" (ISs), and the purely syntactic structures of actual sen-tences are the output of the Sentence Cycle and are called "Externalised Sentences or Phrases" (ESs). In this paper I will only deal with the Lexical Cycle and Internalised Sentences or Phrases, and only to such an extent as is necessary for this study.

Let me now give an example of an Internalised Phrase or conceptual-syntactic struc-ture underlying the lexical item "restaurant" (example taken from Thelen, 1980). It takes the form of

LOC <...>	/WH SB <human> PD <eat> OB <food>/		
where:	LOC	=	Locative, "the place where the eating is done"
	WH	=	category indicating relativisation
	SB	=	Subject, "the one who does the eating"
	PD	=	Predicate, "the action identified with the lexical item"
	OB	=	Direct Object, "the thing that is eaten"
	/_____/	=	these elements can be recovered, that is occur next to the lexical item itself in actual sentences, e.g. "the restaurant where we had dinner"
	<eat>	=	conceptual-semantic component

As is clear from this example, the conceptual-semantic components fill the slots of the conceptual-syntactic categories of this Internalised Phrase. These components are not primitives in the strict sense, but primitives by axiom. The Internalised Phrases or conceptual-syntactic structures of actual lexical items can all be reduced or transformed to one Internalised Sentence. The example Phrase can be reduced to the Internalised Sentence SB <human> PD <eat> OB <food> LOC <...>.

All those lexical items that have at least the conceptual-semantic component <human> in SB position and <eat> in PD position in their ultimate Internalised Sentence are said to belong to one and the same group of lexical items, called LEXICAL SYSTEM. Thus, to the Lexical System of SB <human> PD <eat> ... belong not only "eater", "restaurant", but also "eat", "Yorkshire Pudding", etc. The Lexical System, in other words, is an abstract framework for relating lexical items to one another. Another such framework is the LEXICAL DOMAIN. All those lexical items are said to belong to one and the same Lexical Domain that have one conceptual-semantic component in common in their underlying conceptual-syntactic structure irrespective of the category that is filled by it. The Lexical Domain, in other words, relates Lexical Systems to one another. These two frameworks then are the basis for the comparison of dictionaries in this study. In its initial formulations, TCM claimed to be able to structure the whole theoretical lexicon from nouns to adjectives and prepositions. In the time in between then and now, however, it has been said elsewhere and become clear that the lexicon cannot and should not be structured by means of one and the same principle. It is for this reason, and because in the meantime it has become clear that TCM has problems with adjectives, adverbs and prepositions (that may be structured better by other means) that I decided to restrict myself to nouns and verbs only.

5. The lexical items investigated

Only those lexical items were included that:
   (1) are given in at least one of the dictionaries of investigation;
   (2) belong to the General LS of SB <human> PD <eat> OB <food> and in whose underlying Internalized Sentence or Phrase <human> occupies SB position, <eat> PD position, and <food> OB position;
   (3) are linguistic expressions of the combination of the conceptual-semantic elements "taking into the mouth" and "swallow", i.e. (a) in whose lexical information (or the abstraction of this information) this combination of elements is present, or (b) whose lexical information (or its abstraction) can be "transformed back" to this combination and thus to the General LS of investigation.

For this study I selected the Lexical System of SB <human> PD <eat> OB <food>, which I already partially investigated in Thelen (1980), and the Lexical Domain of <restaurant>. In this study, a slightly adjusted form of the notion of Lexical System will be applied. I will not go into this here. On the basis of the above criteria for inclusion, the following lexical items were selected:

VERBS: LS of SB <human> PD <eat> OB <food>:

"bolt", "break bread", "breakfast", "champ", "chomp", "consume", "cram", "degust", "devour", "diet", "dine", "dine in", "dine off", "dine (up)on", "dine out", "discuss", "dispatch (despatch)", "eat", "eat in", "eat like a horse", "eat like a pig", "eat one's fill", "eat one's head off", "eat oneself sick", "eat out", "eat up", "engorge", "fare", "feast (on/off)", "feed", "feed (up)on", "fill oneself", "fill oneself up with", "fork", "glut oneself", "gobble", "gobble down", "gobble up", "gorge", "gorge on", "gorge oneself", "gorge with", "gormandize", "gulp", "gulp down", "guzzle", "guzzle away", "guzzle down", "guzzle up", "have a feed", "have breakfast", "have dinner", "have lunch", "have supper", "have tea", "ingurgitate", "lunch", "make a pig of oneself", "manducate", "mess", "munch", "nibble", "nibble at", "overeat", "overfeed", "over-gorge", "partake of", "peck at", "pick at", "pig", "pig oneself", "raven", "sample", "scoff", "stuff oneself", "sup", "surfeit", "swallow (down)", "take a meal", "taste", "tuck away", "tuck in", "tuck into", "wolf", "wolf down"

NOUNS: LS of SB <human> PD <eat> OB <food>:

"appetizer/appetiser", "banquet", "bite", "canteen", "consumption", "dining room", "dinner", "fork", "glutton", "helping", "hunger", "knife", "meal", "meat", "messmate", "mouthful", "nourishment", "plate", "refectory", "restaurant", "supper", "table", "vegetarian"

For the LS of SB <human> PD <eat> OB <food>, the total number of lexical items analysed thus was 324 (85 verbs and 23 nouns). This selection is not exhaustive, but had to be restricted for practical reasons. On purpose, also lexical items consisting of more than one item were selected, e.g. phrasal verbs, and prepositional verbs. Per lexical item only three look-ups were carried out. For the LD of <restaurant> only a number of LSs together with a number of member lexical items will be given in this article; they will not be analysed further, simply because the principles of analysis are the same as those for the lexical items of the LS of SB <human> PD <eat> OB <food>.

According to TCM, the following LSs can be part of the LD of<restaurant>: 1) SB <human> PD <eat> OB <food> (LOC<restaurant>) [lexical items: "dine out", "eat out", "menu", "restaurant", etc.], 2) SB <human> PD <serve> OB <food> (LOC<restaurant>) [lexical items: "cuisine", "serve (food)", "wait at table", "wait on people", "waiter", etc.], 3) SB<human> PD <run> OB <restaurant> [lexical items: e.g. "restaurateur"], 4) SB <human> PD <own> OB <restaurant> [lexical items: e.g. "restaurateur"], 5) SB <human> PD <prepare> OB <food> (LOC <restaurant>) [lexical items: e.g. "chef"], 6) SB <human> PD <work> LOC <restaurant> [lexical items: "cook", "waiter", etc.], 7) SB <human> PD <buy> OB <food> (LOC <restaurant>) [lexical items: "bill", "tip", etc.], etc..

In TCM, dictionary definitions are taken to be rough approximations of underlying conceptual-syntactic structures that still need to undergo a process of abstraction. This process entails the analysis and comparison of as many "ordinary" dictionaries (see section (3)), synonym dictionaries and thesauruses. For practical reasons I had to restrict myself to the following synonym dictionaries (1) and thesauruses (2): (1) "Cassell's Modern Guide to Synonyms & Related Words" (Cassell's); "Webster's New Dictionary of

Synonyms" (Webster's), and (2) "Roget's Thesaurus of English Words and Phrases" (Roget's); "Longman Lexicon of Contemporary English" (Longman). The dictionaries (1) and (2) served as references only (in particular for an initial survey of the LS of investigation and for the formulation of inclusion criterion (3) above), and will not be dealt with in this article.

For every lexical item first a Dictionary LS was established and then this Dictionary LS was compared with the General LS of investigation. The General LS of investigation was established on the basis of all the dictionaries of investigation together with all the other dictionaries of reference mentioned above.

6. Discussion of results

For practical reasons only general results can be given here. Detailed results are available on request. For the assessment of ACCESSIBILITY the material was analysed on the following questions: FOR VERBS: 1) how many look-ups are needed to find the form of the lexical items (1 scores best), 2) how many lexical items belong to the LS of investigation, 3) how many definitions are given without synonyms (calculated as percentages of the total number of items given), 4) how many synonyms belong to the LS of investigation (calculated as percentages of the total number of synonyms given), 5) how many look-ups are needed to trace their meaning (1 scores best; result calculated as percentages of total number of synonyms), 6) is the meaning of these synonyms incompatible with the rest of the lexical information and/or the examples given (calculated as percentages of the total number of synonyms with 1 look-up). FOR NOUNS only questions (1) and (2) were answered. For questions 1-2, the scores per dictionary were calculated as percentages of 85 [for verbs] or 23 [for nouns] (total number of items analysed per dictionary). The percentages for the three dictionaries of investigation are for verbs: COLLINS: 57.65 (1), 54.12 (2), 35.62 (3), 34.55 (4), 50.91 (5), 67.86 (6); OALDCE: 38.82 (1), 63.53 (2), 53.62 (3), 28.95 (4), 60.53 (5), 100 (6); COBUILD: 52.94 (1), 70.59 (2), 77.94 (3), 58.82 (4), 41.18 (5), 100 (6). For nouns the percentages are: COLLINS: 100 (1), 73.91 (2); OALDCE: 82.61 (1), 78.26 (2); COBUILD: 95.65 (1), 69.57 (2). If the highest percentages per question are given 3 points, the lowest 1, and the middle ones 2, then the overall number of points per dictionary are: COLLINS: 3 + 1 + 1 + 2 + 2 + 1 + 3 + 2 = 15; OALDCE: 1 + 2 + 2 + 1 + 3 + 3 + 1 + 3 = 16; COBUILD: 2 + 3 + 3 + 3 + 1 + 3 + 2 + 1 = 18. From these numbers, one might conclude (for what it is worth) that Cobuild scores highest on the point of accessibility.

All three dictionaries score extremely well on the point of CONSISTENCY, that is to say, as far as the LS of investigation is concerned. This may be due to the degree of abstraction that had to be carried out on the lexical information for the various lexical items in order to obtain their Dictionary LSs. This may mean either that the notion of LS is not an appropriate measure of consistency, or that the notion itself needs to be reformulated. This requires further research. Unfortunately, the number of nouns analysed is too small to yield any conclusive results for consistence. For this, many more nouns should be analysed.

Because of the different concept of definitions in Cobuild, this dictionary would obviously score very low on EFFICIENCY as formulated for this study. For this reason, it

could not be included here. Apart from this, the same restrictions and shortcomings apply here as for consistency. When analysing the data on efficiency, however, I found that the formulation of this notion, though promising, could not yield any conclusive results, simply because it either is too general or because I did not succeed in translating it into workable parameters.

7. Conclusions

This study has shown that the notions of Lexical System and Lexical Domain can be reliable measures of accessibility of dictionaries (as far as nouns and verbs are concerned), and that they provide useful tools for the analysis of dictionary definitions. They may be helpful in the compilation of dictionaries for specific purposes. The analysis itself has proved to be extremely time consuming. What is needed for a "best possible" analysis of this type is at least a vast number of data and a personal computer with CD/ROM.

Endnotes

Bibliography

ALINEI, M. (1980): "Lexical Grammar and Sentence Grammar. A Two-Cycle Model". In: Quaderni de Semantica. An International Journal of Theoretical and Applied Semantics, 1980/1: 33-95. Bologna: il Mulino.

CASSELL's Modern Guide to Synonyms & Related Words (1971). Ed. by S. I. Hayakawa; revised by P. J. Fletcher. London: Cassell & Company Ltd.

COLLINS Cobuild English Language Dictionary (1987). Ed. by J. Sinclair, et al. London/Glasgow: Collins.

COLLINS English Dictionary (1991). Third Edition. Ed. by J. Sinclair, et al. Glasgow: Harper Collins Publishers.

OXFORD Advanced Learner's Dictionary of Current English (1989). Fourth edition, second impression. Ed. by A. Cowie et al. Oxford: Oxford University Press.

ROGET's Thesaurus of English Words and Phrases (1977). New edition completely revised and modernized. Ed. by R. A. Dutch. London: Longman Group Ltd.

THELEN, M. (1980): "Some Remarks on Alinei's Syntactic Model for the Generation of the Lexicon: the Lexical System of <eating>." In: Quaderni de Semantica, 1980/2: 245-285.

THELEN, M. (1987a): "TCM: A Two-Cycle Model of Grammar. Some Extensions of Alinei's Original 1980 Formulation". In: Aspects of Language. Studies in Honour of Mario Alinei. Volume II: Theoretical and Applied Semantics. Ed. by R. Crespo, B. Dotson Smith and H. Schultink Amsterdam, Editions RODOPI BV, 493-507.

THELEN, M. (1987b): "TCM, A Two-Cycle Model of Grammar, and the Structure of the Lexicon". In: Proceedings of the Fourteenth International Congress of Linguists. Berlin/GDR, August 10-August 15, 1987. Ed. by W. Bahner, J. Schildt and D. Viehweger (eds.) (1989). Berlin: Akademie-Verlag, pp. 1274-1277.

THELEN, M. (1987c): "TCM: een Alternatieve Benadering van de Structuur van het Theoretische Lexicon". In: TTT (Interdisciplinair Tijdschrift voor Taal- & Tekstwetenschap), Vol. 7, Nr. 4. Dordrecht: Foris Publications, pp. 221-237.

THELEN, M. (1990): "From Interpretation to Re-Expression of Meaning in the Translation Process". In: Translation and Meaning, Part 1. Proceedings of the Maastricht Session of the 1990 Maastricht-Lódz Duo Colloquium on "Translation and Meaning", Maastricht (The Netherlands), 4-6 January 1990. Ed. by M. Thelen & B. Lewandowska-Tomaszczyk. Maastricht: Euroterm Maastricht, pp. 288-298.

THELEN, M. (1991): "Categorization and Translation". In: Translation: Theory and Practice; Tension and Interdependence". Ed. by M. Larson. ATA (American Translators' Association) Series, Volume V, pp. 71-86.

THELEN, M. and P. Starren (1991), "Languaging-up the Dictionary Maker: neologisms and dictionaries". To appear in: Proceedings of the IV EURALEX-VOX Congress, Held in Benalmádena, August 28 - September 1, 1990, Barcelona: Biblograf, pp. 116-133.

THELEN, M. (1992, in print): "Stasis and Dynamics of Meaning: The Lexicon in a Two-Cycle Model of Grammar". In: Translation and Meaning, Part 2. Proceedings of the Lódz Session of the 1990 Maastricht-Lódz Duo Colloquium on "Translation and Meaning", Lódz (Poland), 20-22 September 1990. Ed. by B. Lewandowska-Tomaszczyk and M. Thelen.

WEBSTER's New Dictionary of Synonyms (1973). Ed. by B. Gove, et al. Springfield, Mass.: G. & C. Merriam Company, Publishers.