

Pike

Submitted to
Anthropological
Linguistics

Oct. 1964

Semantic Fields and the Mansaka Thesaurus¹

David and Dorothy Thomas

Summer Institute of Linguistics

University of North Dakota

1. Thesaurus concepts
2. Classification from dictionary listings
3. Fields and homonymity
4. Zero lexemes
5. Outline of the Mansaka Thesaurus
6. References cited

July 4, 1964

Semantic Fields and the Manskala Thesaurus

David D. ^{the} Thomas
Summer Institute of Linguistics
University of North Dakota

1. Thesaurus concepts
2. Classification from dictionary listings
3. Fields and homonymity
4. Zero lexemes
5. Outline of the Manskala Thesaurus
6. References cited.

1. Thesaurus concepts.

Classification of words in terms of their form (i.e. a standard dictionary) is relatively simple because the number of phonemes in any language is small and there is also a fairly standardized alphabetical order. But a classification in terms of meaning is far more difficult because of the thousands of lexemes involved, and also because there is no standard grid on which they can be placed and classified. A Classification of meanings should include idioms as well as single words.

Though difficult, a grouping of related meanings has considerable practical value, as when attempting to express thoughts with precision in one's own language or in another language. Several approaches to the problem have been suggested or attempted.

Roget's Thesaurus attempted the standardization of a universal grid based on a logical a priori subdividing of general concepts. Roget had hopes that this would help to prepare the way for the construction of an international language. This approach would

result in an etic rather than an emic classification if used as an approach to other languages. (Cf. Conklin 1962 on this point.) And as Pike (1964) points out, construction of useful etic grids, like the standard phonetic charts, should result from the analysis and comparison of many different languages, rather than being constructed a priori and then used to make analyses.

Another approach, advocated by Ullmann (1953), attempts to classify words on a basis of motivation, generic terms, synonymy, homonymy, emotive devices, etc. This approach, however, seems better for typological comparisons than for practical language description.

A third approach, suggested first by Trier (1931) and elaborated more recently by Conklin (1962) and Frake (1961), divides meanings by a taxonomic hierarchy of inclusiveness. Studies of this sort have usually limited themselves to single fields such as the terms for knowledge, color, persons, diseases, plants.

A fourth approach is that of the situational domain (field), proposed by Perzig (1934), in which all terms bearing on a specific cultural activity or situation are classed together. E.g. 'going' is an activity of the feet, so would be classified with feet, and 'blonds' would be classified with hair, and cures would be classified with diseases. An approach somewhat

similar to this was suggested by Voegelin and Voegelin (1957) in classifying about 1500 Hopi words into domains, though with somewhat more conservative groupings than Perzig. Cooccurrence in text (Voegelin p.4) or cooccurrence in conversation (Frake p.) gives useful domain groupings. The Mansaka Thesaurus was compiled from dictionary listings, using the principles described in Sec.2, as not until ~~the~~ near the end did we get even sporadic informant help.

The approach followed in the 4000-word Mansaka thesaurus is essentially that of modified situational domains, though, like the Voegelins, ^{we} found taxonomic hierarchies more practical at some points than situational domains. Fishing, for instance, was handled as a situational domain, while Fish was handled as a taxonomic hierarchy. And the practical exigencies of economical classification dictated the placing of Fish and Fishing in unrelated sections of the thesaurus. These complications were almost unavoidable, because of practical complex interrelations; e.g. fish are relevant both to fishing and to food. This is a weakness which can be overcome by making cross-references between mutually-relevant sections.

Ullmann (1953 p.227) felt that the taxonomic field approach could not be used for a total language description. It would

indeed seem that a complementary use of taxonomies and domains might be best, as taxonomies are more satisfactory for discrete similar items, and domains are more satisfactory for complex activities. Frake's discussion of Subanon diseases is essentially several unrelated hierarchies brought together in one domain.

"Field" as used in this combined way can be defined generally as a terminological set the members of which have very similar sets of possible cooccurrences. If held to a single grammatical class, this definition would yield taxonomies, if broadened to allow all classes it would yield domains. Throughout the rest of ~~this~~ this paper this is the concept referred to as ~~field~~ a "field".

Katz and Fodor (1963) dismissed culturally-oriented semantic descriptions as impractical, but the recent studies by Conklin, Frake, Franklin, etc., have ~~definitely~~ clearly demonstrated their possibility and their practicality.

2. Classification from dictionary listings.

The problem of the classification of meanings is touchy. As I. A. Richards (1931) commented, ''To criticize Roget's categories would be to bring up all the hardest problems there are.'' The Mansaka Thesaurus was prepared largely from dictionary file slips² with relatively little benefit of informant or text, but the following principles were found helpful and gave what I feel is a fairly emic picture of Mansaka domains.

a) The classifications π should be set up on the basis of semantic rather than grammatical criteria. Thus nouns, verbs, adjectives, etc., can often be mixed together in a single domain. Not infrequently, however, grammatical and semantic divisions do coincide. (Strict taxonomies are generally restricted to a single word class.)

b) Major fields filled mostly with affixed forms or phrases rather than with unaffixed root morphemes should be regarded as suspect as a major grouping. Thus Colors and Shapes were highly suspect as valid major fields in Mansaka because they were mostly affixed forms. (But cf. Conklin 1962.) Minor fields or subfields, or fields

involving cultural innovations, can often be composed of compound or affixed forms. All unitary lexemes, whether simple or complex (words or phrases), should be included in the thesaurus.

c) Forms containing active affixes should belong usually to the same general field as the root. Thus loto 'cook', lotoanan 'stove, oven', pagloto 'cooking', maglotoway 'one who cooks', etc., all fit in the same general domain of Cooking. Forms containing fossilized affixes, however, may have shifted into a different domain from the root.

d) Fields should usually be describable by a single word in the language; thus Panaw 'Traveling', and Panday 'Craftsmanship'. But a complicated label like Yagakanaguna nang Otay 'Attitudes of People' was suspect and was eventually replaced by the single word Ontol 'Character'. Semantic fields involving grammatical connectives, ligatures, and the like, may be difficult to find simple names for; and it is my feeling now that perhaps grammatical function words belong to a different universe, not just a different field, and they should not (can not?) be classified in

terms of lexical fields, but belong instead in the grammar alone. (The Voegelins 1957 hoped to include grammatical terms in their domains, but hadn't yet been able to do so.)

e) Division of natural phenomena should follow local (emic) classifications rather than scientific classifications. Thus the Mansaka vegetable kingdom was divided into Trees, Vines, Plants, and Grasses.

f) When there is extensive cross-classification of categories the arrangement should be carefully reexamined. Thus an early domain of Articles and Utensils was at cross-purposes with such domains as Fishing, Agriculture, and Craftsmanship, and it was eventually abandoned because the other posited domains proved sufficient. (Conklin 1962 points out that a certain amount of cross-classification should be expected, but for practical purposes this should be held to a minimum.)

g) Some fields will tend to be more closely integrated and definable than others. Fields such as pronouns, colors, etc., are more often closely integrated, with

fairly clear componential analysis possible. As a result these are more often chosen for special semantic analysis. Other fields tend to be more difficult to circumscribe, nevertheless they should be considered equally real.

Some fields may be held together by interlocking components, as is the case with many pronoun systems, others by relationship to a central nuclear concept, e.g. disease (Frake 1960); others by a serial relation, as with numbers; others by mutual complementation within a definite circumscribed area of meaning, e.g. botanical terms, colors, body parts.

Even as the physical distinction between *pr* phonemes is sometimes blurred at the edges, and distinctions between clear grammatical classes sometimes get indistinct at peripheral points, so the divisions between semantic fields will often be indeterminate at certain points.

Some fields will be tightly closed systems, as with numbers in most languages and pronouns in some languages. Other fields will be quite open and fluid,

as with animals, or clothing, or agriculture in many languages. I believe that this has a bearing on the concept of basic vocabulary, as closed systems, along with tangible discrete units and daily usefulness, seem to have a high degree of morpheme persistence. (Cf. Kroeber 1963.)

1) In moderation of the above principles, it should be expected that there will be indeterminacies, a number of dual entries, and even some contradictions, from the natural indeterminacies of language#.

ix 3. Fields and homonymity.

We assume that concepts and percepts/^{that}are functional in a culture will normally be distinguished in the language (with perhaps a time lag between culture changes and corresponding language changes).³

It would follow from this that terminological groupings within a field are not just an arbitrary linguistic representation with greater or lesser correspondence with cultural reality, but that the lexemic groupings may be assumed to correspond fairly accurately with the ~~present~~ present or a slightly previous structuring of that field in the culture.

A second corollary is that a language will not long tolerate ambiguity where it is culturally misleading, or where it does not correspond with cultural ambiguity. ~~xxxx~~ Homophony usually occurs between members of different fields. E.g., Eng. pear and pair belong to two unrelated fields, those of 'fruit' and 'quantity', and there is very little chance of ambiguity (that is, the set of non-trivial contexts in which the one occurs is quite different from the set of such contexts in which the other occurs.)

Where there is ambiguity (homophony) in terminology within one field it may then be assumed to reflect some cultural ambiguity. This normally seems to occur in two types of situations:

a) When the ambiguity occurs between continuous⁴ sections of a field, the two ~~xxxx~~ homophonous morphs may be considered to be the same lexeme. E.g. Eng. house (large) and house (small)

may be considered to be continuous etic percepts in the field of Dwellings, so the two uses of house would be a single lexeme and presumably a single emic concept. (This seems obvious to native speakers of English, but might not seem so obvious to an observer whose native language clearly distinguished large houses from small houses.)

b) When the ambiguity occurs between a whole and a fully-included part of that whole, the whole may most frequently be considered an extension of the meaning of the part, or sometimes the part may be considered a specialisation of the meaning of the whole. Thus, Eng. man 'male humans' and men 'humans'; Span. padre 'father', padres 'parents'.
They would be separate but related lexemes.

Procedurally this principle would indicate that when there is ambiguity between two seemingly non-continuous parts of a field, the analyst should carefully recheck his data to see if he has mistaken the emic structure of the field and that a reanalysis would show the parts to be continuous. For instance, kinship terms not infrequently have etically distant relationships bearing the same kin term; emic analysis should generally be expected to reveal such relationships as emically continuous.

Similarly with a part and its whole, if the denotation of the part is not fully contained in the whole, the analysis should be suspect of being incomplete. E.g. man 'human' should be expected to entirely include all man 'male humans'.

The term homonym should perhaps be reserved for non-continuous forms from different fields, with items in a part/whole relationship being considered extensions of or specialisations of a same lexeme. Thus pair and pear would be true homonyms. Man (vs. animal) would be an extension of man (vs. woman). Man (vs. unmanly male) would be a specialisation of man (vs. woman). Man (as in 'man the cars') could perhaps be considered an extension of man into another field, though this might be subject to some question.

It might seem also that when there is part/whole ambiguity, the part involved will be that part which is most predominant or which is most significant culturally. Thus the extension to man (vs. animal) is not surprising in our basically patrilineal society, and it is also not surprising that the generic man, man, seems to be losing out to people, in the light of our bilateral kinreds and our equalitarian tendencies.

As for the question of basic vs. derived meanings, an attempt at a clear procedural statement would be premature, but one or two remarks may be made. Historical derivation and current derivation are not necessarily the same. Current derivation may sometimes be ascertained by seeing which field associations or opposites come most readily to mind when the word is mentioned. Though historically related, meanings that are currently non-continuous must be considered separate lexemes.

Katz and Fodor make no contrast between basic and derived meanings, being computer-oriented rather than culture-oriented. The non-recognition of semantic fields led them also into the error of making their primary division of meanings on the basis of grammar rather than meaning, producing, as in their example of play, separation of related meanings and joining of unrelated meanings.

4. Zero Lexemes.

Goodenough (1956) postulated a zero lexeme in Truk kinship. A culturally significant category existed for which there was no specific lexeme. Zeroes have been postulated in morphemic analysis when there is a significant gap in a paradigm, so a lexemic zero would fit the analogy.

We should make clear the distinction, however, between cultural zeroes, language (lexemic) zeroes, and culture-language correlation gaps. A cultural zero would be a specific abstinence from or avoidance of an act which might have been expected to fill out a cultural matrix or pattern. Taboos and avoidance patterns might be among the clearest examples of significant cultural zeroes.

A language zero would be the specific absence of an overt form for a systemically expected language form or an expected combination of matrix components. The absence of all other forms in a particular position implies the presence of this specific category. This is of course the morphemic zero principle, which would apply also to lexemic zeroes. Lexemes (as described by Conklin) are simply the next level above the morpheme in the morpho-syntactic hierarchy, and in the majority of cases the lexeme and morpheme are identical.

Different again from this is a culture-language gap. This would be the situation where there is a culturally significant category or act for which there is no specific term. On our previous assumption (sec.3), language will usually reflect culture. But this category would take care of situations where, perhaps because of novelty or some other such reason, the language has not yet codified a specific term. This is not a cultural zero because the cultural category actually exists, and it is not a lexical zero because the category in question is not automatically understood from the absence of all overt fillers from an expected grammatical position. This culture-language gap would seem to be the category that best fits Goodenough's definition, but I would consider it a gap rather than a significant zero.

None of these three types of categories were used in the Mansaka Thesaurus, as I know of no linguistic zeroes in Mansaka, and my cultural analysis hadn't progressed far enough for me to perceive cultural zeroes or culture-language gaps.

It may be objected that there is a difference between a zero in a field or matrix (lexical) and a zero in a sequence (grammatical). But my position on this would be that a grammatical zero is just a special case of a lexical zero where the field involved is co-extensive with a total grammatical category. Grammatical (morphemic) zeroes have usually been posited in situations where the grammatical category involved is composed of just a single restricted lexical field, as with pronouns in many languages. And in turn, a ^{taxonomic} lexical field can be considered a restriction of a grammatical category to a specific cultural domain or set of co-occurrences in the same discourse setting. So that essentially the same principles could be used for both.

A third corollary to our previous assumption (sec.3) would be that if zero lexemes are posited, there can be only one such zero in any one field, as the two zeroes would be homophonous, causing culturally-unnecessary ambiguity.

5. Outline of the Manksa Thesaurus.

I. DUNYA - the world

A. Langit - the sky.....heavenly bodies, time, clouds

B. Bansa - the earth.....

1. Bansa (general).....weather, heat, light

2. Lapa - land.....dust, stones, hills, geographic features, metals

3. Taling - water.....sea, depth, floods, pouring, whirlpools, swamps,
submerging, fresh water

C. Kinaba - time.....duration, speed, frequency, recentness, when

D. Kyakabutangan - relative location..here, below, above, behind,
direction, nearness

E. Bilang - number

1. Sakal - measure.....small, large, measurement, exactness, difference

2. Gundang - estimate..amount, shapes

3. Kadag - many.....many, all, increase

4. Tapang - volume

5. Tihang - weight

6. Bilang - counting...cardinals, ordinals

II. MANAHAP - animal world

A. Langan - birds.....bird characteristics, chickens, hummingbirds,
sparrows, woodpeckers, parrots, doves, hornbills,
fishers, bats, owls, eagles, unidentified

B. Manahap - animals.....

1. Manahap - animals...gnawing, grazing, animal characteristics, carabao,
goats, pigs, monkeys, dogs, lizards, snakes

2. Manung - insects....lice, mosquitoes, ants, flying insects, bugs,
beetles, worms, termites, snails, spiders

Kyakabutangan

C. Ida - fish

1. Ida nang Tobig - freshwater fish
2. Ida nang Dagat - saltwater fish
3. Kasili - eels
4. Orasang - shrimp
5. Kagsang - crabs
6. Kabasan - shellfish..clams, snails

III. GIMIGITI - Plant world

- A. Gimigiti - general terms..leaf, root, ripe, decay
- B. Tanum - plants, crops..bananas, coconut, bamboo, vegetables, fruit
- C. Kasy - trees.....tree parts, woodcutting, misc.trees, palms
- D. Baragun - vines.....
- E. Sagbut - grasses.....grasses, ferns, weeds

IV. OTAW - Man

A. Lawas - body

1. Napas - life
2. Lawas - body.....bodyblood, skeleton, skin, appearance
 - a. Oro - head.....hair, face, eyes, nose, cheek, mouth
 - b. Lawas - body.....chest, stomach, internal organs, pelvis
 - c. Bakun - arms
 - d. Siki - legs
3. (postures).....bow, stand, sit, embrace
4. Ipa - babies.....pregnant, birth, babies
5. Paghita - seeing....curious, look upon, search for
6. Pagtanggap - hearing..hear, noisy, sounds

7. Pagtas - smelling....smell, odors
8. Kaanag - strength....strength, weakness, fatigue
9. Sakit - sickness.....pain, dissiness, colds, fevers, aches, fractures
 - a. Paris - skin diseases..bites, boils, burns, itches, wounds
 - b. Oro - head diseases
 - c. Lawas - diseases of the body

B. Ontol - characteristics

1. Ontol - character...manner, similarity, influence, diligence, carefulness, confusion, hope, fear, peace, shy, shame, evil, trouble, abuse, greed, dishonesty, impatience, suitability, envy, pride, love, anger, happiness
2. Oual - address.....age, status, lineage, reputation, pronouns
3. Layon - relatives...blood relatives, in-laws
4. Batasan - customs...legends, betel, dancing, music, weddings, burial, visiting
5. Olan - authority
6. Tinabangay - cooperation..help, accompany, agree
7. Pagbatak - marking..color, writing

C. Inang - motions

1. Inang - motions.....reach, press, move, glue, pretend, wander
2. Pagillayon - communication
 - a. Saceritin - speech..voices, talking, question, command, slander, shout, whisper
3. Bagan - mind.....think, understand, believe, learn, plan, prefer

4. Pagkarogso aw Pagkalleg - falling, leaning
5. Pagbarin - turning..go ahead, turn, turn back
6. Pagbutang - putting..place, insert, remove, pull, hang, scatter
7. Pagkapeti aw Pagkasogat.- touching, hitting..touch, strike,
punch, tap, squeeze, throw
8. Tagun aw Borey - holding, releasing..hold, keep, release, lose

D. Imo - doings

1. Imo - doings(misc.)..deed, try, able, bend, mix, own
2. Potes - bundles.....packages, wrapping
3. Pagkasapad - breaking..destroy, tear, cut, grind, burst, pierce
!
!xxxBontenxxxptine
4. Pagtabon - concealing..cover, hide, search, open, display
5. Bonton - piles
6. Kamang - getting.....get, grab, pick up
7. Ocos - tying.....tie, loop, twist, tighten, break loose
8. Pagtabaray - finish..accomplish, use up, stop, begin
9. Paglinis - cleaning..clean, dirty, weeds, sweep, wash
10. Paglasak - contain..fill, baskets, bags, boxes, tubes
11. Longag - holes

E. Uya - living

1. Uya - life.....relax, visit
2. Baray - housing.....shelters, neighbors, furniture, ladders,
floors, rooms, roofs

3. Pananapuan - clothing..cloth, laundry, sewing, weaving, dressing,

jewelry, clothes, head coverings, fit

4. Korang - sleeping...sleepy, beds, lying down, sleep, waking

5. Atoron - fire.....ignite, burn, smoke, ashes, charcoal, lights

6. Tarapag - provisions for cooking..supplies, seasoning

7. Pagloto - cooking...stir, pots, preparation and preservation of food,

cooking methods

8. Pagkaan - eating....hunger, taste, eating, food, drink

9. Ballung - play.....toys, contests, aimless play, cock fights

F. Gawbak - work

1. Gawbak - work(misc.)..work, use, work hard, guard, knives

2. Karindong - transactions

a. Pantiyari - selling..business, money, sell, prices, debts

b. Inatagay - barter.give, exchange, haggle

3. Agos - hunting.....snares, traps, trails, killing

4. Pallopa - agriculture..prepare field, plant, sprout, harvest,

abaca, rice, winnow, sprinkle, carabeo work

5. Gakaya - fishing....nets, traps, hook and line, harpoon, poison

6. Panday - craftsmanship..materials, tools, tool parts

7. Panaw - travelgo, around, through, guide, cross, provisions,

meet, depart, arrive, land travel, boats, ascend

8. Tagedara - carrying..loads, carry, drag

9. Digun - strengthening..weak, flexible, strong

10. Pagtanawan - warfare..war, stalk, attack, defend, conquer

V. KANANG DIWATA - religion...spirits, ceremonies, punishment

VI. DOGTONG na SORIT - language connectors...if, because, but, when, how,
surely, perhaps, even, focus markers, yes, no,
concerning

6. References cited.

Conklin, Harold C., 'Lexicographical treatment of folk taxonomies', in Householder & Saporta eds., Problems in Lexicography (IJAL Suppl. 1962), pp.119-141.

Frake, Charles O., 'The diagnosis of disease among the Subanon of Mindanao', Am.Anth 63:113-132 (1961).

Goodenough, Ward H., 'Componential analysis and the study of meaning', Lg. 32:195-216 (1956).

Kroeber, A.L., Yokuts Dialect Survey; U.Calif.Anthropological Records 11:3 (1963).

Katz, J.J. and J.A.Fodor, 'The structure of a semantic theory', Lg. 39:170-210 (1963).

Ohman, Suzanne, 'Theories of the linguistic field', Word 9:123-134 (1953).

Pike, Kenneth L., 'Discourse analysis and tagmeme matrices', to appear in Oceanic Linguistics supplement 1964.

~~Bornig, 'Wissenschaftliche Bedeutungsbeziehungen', Beit.z. Geschichte d. deutschen Sprache u. Literatur 48: 97-109 (1934).~~

Richards, I.A., Introduction to Rogot's Thesaurus 1951 ed.

~~Gries, Jost, Der deutsche Wortschatz im Sinnesbereich des Verstandes: die Geschichte eines sprachlichen Feldes (1931)~~

Ullmann, Stephen, 'Descriptive semantics and linguistic typology', Word 9:225-240 (1953).

Voegelin, Charles F. and Florence M., Hopi Domains, IJAL Memoir No.14 (1957).

Footnotes

1. Mansaka is a Malayo-Polynesian language spoken in Davao province in the Philippines. The Mansaka Thesaurus was prepared in 1956-57, with 3 typescript copies, one each in the possession of Gordon Svelmoe and the Philippine and Vietnam libraries of the Summer Institute of Linguistics.

The first author did the larger share of the writing of this paper, the second author did the larger share of the actual work on the thesaurus. We are indebted to Ernest Lee and Alan Healey, among others, for helpful comments on an earlier draft of this paper.

2. File slips with illustrative sentences, prepared by Juan Flavia and Melanio Dupla, native speakers of Mansaka, under the guidance of Gordon Svelmoe and Norman Abrams, of the Summer Institute of Linguistics.

3. Cf. Murdock's approval of Lowie's statement that 'language represents reality and ... in so far as it is related to social phenomena it is likely to mirror them.' G.P. Murdock, Social Structure, p.118.

4. Voegelin & Voegelin, 1957, p.4ff.