A Construction Theory of Thematic Map Taking A Carto-Linguistic Perspective

Fei Zhao*, Qingyun Du, Cong Wang*, Jianjun Liu***

* National Geomatics Center of China, Beijing,China
** Wuhan University, Wuhan, China

Abstract. The construction for thematic map symbol is a very complex and intelligent process. This symbol can be automatically generated and easily shared on the web through the syntactic structure of semantic symbols. The symbol types, inner structure and design pattern are expounded. And a syntactic construction theory based on letter (thematic maps primitive) - word (single thematic symbol) - sentence (combined symbols or complex symbols) structure model is put forward for automatic construction of thematic map symbol. As a result of this research, symbols can be defined using cartographic primitives which are arranged according to its syntactic principles. Then the semiotic model and word-centered construction theory can be integrated into interactive cartography represented by the technology of Internet. Finally, its concept and schema is discussed, and some examples are presented based a web thematic cartographic system to verify its power.

Keywords: Syntax, Thematic map, Symbol, Linguistics

1. Introduction

With the development of Internet of Things (IOT), thematic cartography is required to be real-time and intelligent by the dynamic monitoring of sensor data(Iosifescu-Enescu et al., 2010). Intelligent thematic cartography is in the ascendant. It develops into intelligent selection of expression method(Jing et al., 2007) instead of determining the expression content based on expert system(Yixin, 1993). Now intelligent thematic cartography is focusing on dynamic intelligent generation of thematic symbol. There are two approaches: one is to analyze elements of symbols(Yaofeng, 1997),

Schnabel & Hurni, 2009), and then dynamically construct symbols according to constitutive rules of given symbols; the other is to mining thematic mapping knowledge from the produced atlas(Hongsheng et al., 2009), and then select thematic symbols based on case reasoning mechanism of artificial intelligence. The former method is to generate symbols as a result of elements and constitutive rules, it can extend thematic symbol types. And its defect is that the incompletely constitutive elements and rules may lead that produced thematic symbols do not conform to spatial cognition rules. The latter method can produce thematic maps which are similar as cases according to existing and excellent thematic maps or atlas. However, it cannot satisfy with dynamic mapping in a distributed environment, in that it depends on existing symbol characters. Web2.0 users have much more interactive means(Boulos et al., 2010), and more incline to participatory spatial information service(Jankowski, 2009). How to analyze cartographic data rapidly? How to choose appropriate representation mode? How to design the corresponding symbols? Which construction modes should be followed in order to generate dynamic thematic maps? The questions above are to be solved.

This paper aims to introduce semiotics and linguistics into thematic cartography, analyze map symbol system from the perspective of phonetics, semantics, syntax, abstract semiotic model of thematic maps, and construct thematic symbolic framework centered on “word“. These help define thematic symbols according to three attributes which are symbolic elements, structure principle and word relation, and select constructing attribute intelligently based on different data characteristics, and then realize to generate thematic maps automatically. Discussing design model of this mechanism, and giving formal describing of prototype system and its integrated application.

2. Syntactic Model of Thematic Map Symbol

The concern with linguistics of cartography can be divided into macroscopic and microcosmic level. Microcosmic linguistics mainly focuses on symbol structure elements, meanings, relations inner symbols and layers, and construction mechanism(Qingyun, 2004). According to this, map symbol system can be divided into following parts which are from elementary to advanced: ①Symbol construction elements(Primitive, Visual Variable, etc.) ——Phoneme；②Single thematic symbol——word；③combined or complex symbols——sentence. Phonemes as minimal phonetic unit form word according to semantic structure, words form complex symbols

through syntax, and further form symbol layer. And symbol layer can form thematic maps through overlying with other layers. “Phoneme” should form map symbol system that has syntactic structure in a unified semiotic meaning framework. The model is three-dimension space that includes symbolic primitive (phonetic), symbolic reference (semantic) and symbolic organization (Syntactic) . (Figure 1)

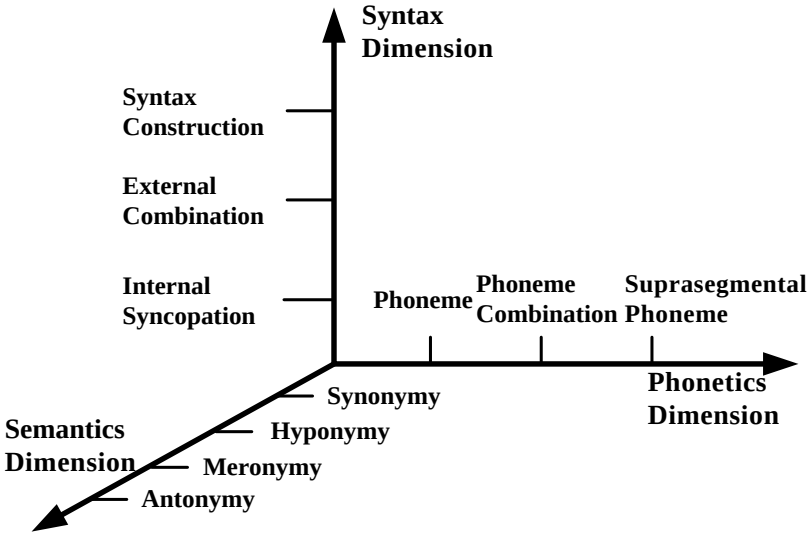


Figure 1. Syntactic model of thematic map symbol

Phonetics-dimension is divided into three levels based on their different degrees of abstraction: ①Phoneme is the most basic distinctive feature among symbolic constitution elements. Point and line which are controlled by six distinctive features can for symbolic phoneme. And the six distinctive features include size, shape, orientation, brightness, color and saturation. ② Phoneme Combination: Phoneme can form “Word Morpheme” (primitives) and “None Word Morpheme” (visual variables); ③ Suprasegmental Phoneme is combination phonetic feature. It described symbol element in detail such as area graphic line of curvature. Semantics-dimension includes symbolic referential features and semantic features. Thematic symbol system focuses on semantic relations of words that are synonymy, hyponymy, meronymy and antonymy. Syntax-dimension has three levels: internal syncopation of words, external combination and syntax construction. Internal syncopation of words means phoneme set; External combination means words form phrases based on the semantic relations; Syntax construction consists two meanings, one is repeated using

structural principle between words or words and phoneme set based on semantics to form symbolic sentence which is a kind of combined relationship; the other is that secondary phoneme of sentence level represents new feature.

Phonetics-dimension means define phoneme of symbols, forming morpheme, and then forming words. The construction of this dimension is controlled by inherent constructive feature of symbolic phoneme. And it cannot consider the semantics and syntax. Semantics-dimension and syntax-dimension have effect on morpheme word formation and word combination into sentences. The common carrier of syntactic structure is word. Word is the most active part in the symbol system. And it is both the components in abstract language system and a specific unit in real symbol system. Through the investigation of this kind of "active symbol", researchers can hold the word-formation regular rules of thematic symbols, and can further combine with specific data and context to research pragmatic process of word meaning. Therefore, it is important to make the word to the core position, and form the word-centered model for thematic symbolization framework (Fig.2). The framework includes three macro levels that are phonetic, semantics and syntax, corresponding with syntactic structure. Word and sentence are two material carriers of symbols. Three macro levels have comprehensive effects on the process of word construction, and then combine to generate sentence.

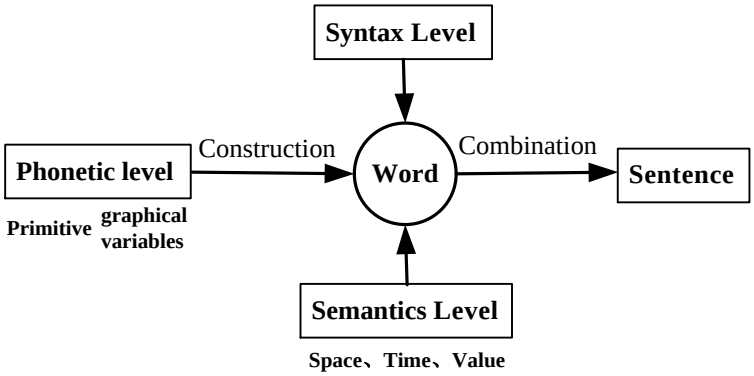


Figure 2. Word-centered model for thematic symbolization

3. Thematic Symbol Construction Model based on Syntactic Structure

According to syntactic structure of thematic symbols and symbolization framework in this paper, combined with design principles of thematic symbols, made the Primitives (*Primitives*, *P*) which hava nature construction property, retinal variables (*Visual variables*, *V*) mentioned above as basic symbol elements, forming phoneme set (*Symbol Phoneme*, S_p). Operator set (*Structure Operator*, O_s) which employed symbol structure principles constructs word set (*Symbol Word*, S_w) based on phoneme according to the structure of the symbol building operators. Finally,sentence set(*Symbol Sentence*, S_s) including combined or complex symbols is generated by the combination between phoneme and word based on the Structure Operator.Therefore, Symbol set(*Symbol*, *S*) can be described as followed:

$$S = (S_p, S_w, S_s)$$

And in which, $S_p = O_s(P, V)$ $S_w = O_s(S_p, S_p)$ $S_s = O_s(S_p, S_w)$

This part can be divided into three basic aspects: the definition of symbols phoneme collection (S_p), construction of symbols word collection (S_w) and combination of symbols sentence collection(S_s).

3.1. Symbol Phoneme and its Combination

As mentioned above, symbol phoneme is the integration of basic primitives and visual variables. And its combination can form all morphemes(minimal meaningful unit) or rich syllables(still meaningless unit) in phonetics level based on certain rule, such as Aesthetics, gestalt law and so on. It must extract morpheme level symbolic elements to study the characteristics of word-building ability themselves from phoneme combinations which also have a great deal of flexibility.So this section is aimed at determining primitives and the corresponding visual variables with different word-building ability to form phoneme set.

There are different views on the basic primitive type(Xiaoning, 1986), Wei & Yongqian, 1997), Schnabel & Hurni, 2009). The former two aim to generate thematic symbols with single combination, but neglect the mapping relation between data and primitive or symbol. Although the latter represent thematic data through the control of the scale of statistics symbol data direction,but it limits to ideographical expression at carto-word level regardless of intelligent combination at syntactic level. Considering these

facts, this paper extracts thematic symbol morphemes from phoneme combination forming primitives and analyzes their cartographic characteristics, such as position, structure, data mapping and so on.

And the disjunctive principles are as followed:

- It must be generated from carto-phoneme through limited regular steps with a clear reference point;
- There must be a proper graphic scale with construction properties to constitute a continuous structure;
- There is up only one dimension of geometric properties to participate in data mapping and other dimensions directly using visual variables to control.

Based above and with reference of Primitive-based Construction Theory(Schnabel & Hurni, 2009),the *Symbol Phoneme set* in this paper can be abstracted as the phoneme and its basic combination or three-dimensional form showed in Figure 3. The special construction properties marked in red colour in Figure 3 include reference point,which indicating the position in the map,and size or orientation of different dimensions.The phoneme primitives above already have a certain visual variable properties in the level of *Symbol Phoneme set*, such as outline width or color, filling color or pattern,etc..According to the construction properties, primitives can represent each dimension of thematic data by deflection change,size scaling, etc..But the semanteme of symbols or thematic data contents can not be realized in phoneme level.It must do further discussion on the construction principles in word level.

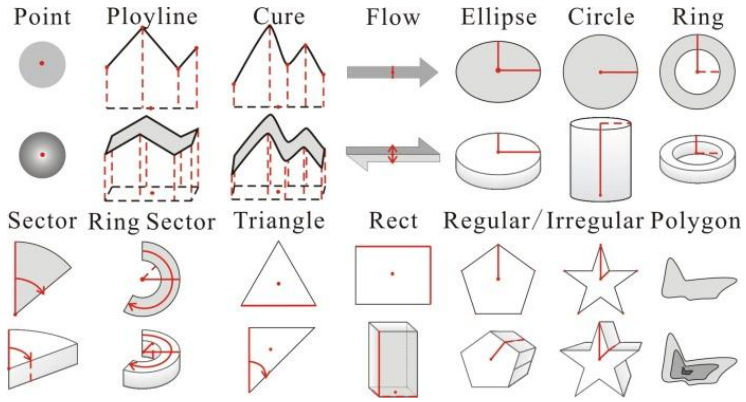


Fig.3 phoneme and its combination of thematic map symbol

3.2. Construction Principles in Word Level

On the basis of analysis of the *Symbol Phoneme set* construction properties, *Symbol Word* (S_w) can be constructed with S_p following below *Structure Operator*(O_s). Here we borrow ideas from current diagram construction theory (Bertin, 1983), Wilkinson & Wills, 2005), Schnabel & Hurni, 2009) and extend it.As showed in Figure 4,the black solid line indicates the data mapping direction of S_p and the dashed one indicates the construction direction:

1. Point: Each s_{pi} in S_p constructs word with itself located by the reference point.The construction property which to represent thematic data usually is mapped with a statistical indicator (*Indicator*, I_i)by Size, such as height of bar symbol,radius of circle, side length of regular polygon andso on.And it is usually used for proportional symbols binding none word morpheme(shape, color etc.)(Figure 4a). Furthermore, it can be used with correlative construction principles integrally by addition of a series of *Operators*(Rotation, Reflection etc.).It can be described as followed:

$O_s.Point(s_{pi}.position,s_{pi}.size(I_i),s_{pi}.color.Hue(I_i),s_{pi}.rotation,s_{pi}.reflection)$.

2. Linear:Multiple $s_{pi}(s_{p1},\ldots, s_{pn})$ located by the reference point in S_p construct word along linear direction with specific interval. It is used in the visualization of two or more dimensional data values,such as bar, stacked bar, dynamic circle and line chart etc..According to the relation between S_p construction direction and data mapping direction, it can be divided into there different concrete modes: parallel, orthogonal and angled(correspond to Figure 4b, c and d).This construction principle has a strong linear directivity so that it cab be used when the s_{pi} has regular pattern in a certain direction.Here we denote the relation type of the direction between construction and data mapping by Linear.direction, the interval of s_{pi} by Linear.interval.So it can be described as followed:

Os.Linear(Linear.direction,Linear.interval…,Os.Point(spi.position,spi.size(Ii), spi.color.Hue(Ii)),…).

3.Polar: Multiple $s_{pi}(s_{p1},\ldots, s_{pn})$ located by the reference point in S_p construct word along polar direction with specific rotation angle.The s_{pi} with polar coordinates properties can directly constructed by this principle based on their center.Others need auxiliary elements(a circle generally) to make themselves have above properties. It is used in the visualization of two or more dimensional data values,such as circle, ring, sector and wing chart etc..In this principle,the S_p construction direction is polar, meanwhile,the data mapping direction can be divided into: polar and circular elevation (correspond to Figure 4e and f).Here we denote the region of thematic data by Polar.direction, the value of s_{pi} by

Region.value and the num of s_{pi} by Region. num.Then the num of each row or column of s_{pi} and the distance themselves can be calculated by Region.area and Region.value.It can be described as followed:

Os.Polar(Polar.direction,Polar.startAngle,...,Os.Point(spi. position,spi .size (Ii), spi .color. Hue (Ii)),...,Polar.angle, Polar.refCircle)。

4. Region：Multiple s_{pi}(s_{p1},..., s_{pn}) in S_p construct word in a certain space(Grid, Region etc.) repeated permutation by row-column(Figure 4g).Each s_{pi} represents a certain equivalent value and the total number of primitives represents the visualized data value so that it is an important means for compensating the shortage of quantitative expression of other symbols. Several groups represented each different value can separately constructed by Region then by Line(using color todistinguish) to represent multi-dimensional thematic data,such as coin charts, dot chart(the space is thematic region).Here we denote the region of thematic data by Region.area(Grid or Polygon), the value of s_{pi} by Region.value and the num of s_{pi} by Region. num.Then the num of each row or column of s_{pi} and the distance themselves can be calculated by Region.area and Region.value.So it can be described as followed:

Os.Region (Region.area, Region.value, Region.num, Os.Point (spi. position, spi .size (Ii), spi .color. Hue (Ii))).

5. Topological： Certain allophones and word morpheme in S_p construct based on topological relation. It will mainly represent cartogram, isogram, choropleth and flow. In cartogram, general topological relation is unchangeable in S_p, but to change the geometry of S_p, it will realize representation between thematic data and topological distance or topological area; Interpolating contour based on discrete S_p objects to generate isogram; in order to get isogram, we should use thematic data which are represented by polygon phoneme itself to match different color or mode after classification; flow is aimed at motion property of thematic phenomenon to represent flow condition between two sites. It is constructed by selecting unidirectional or bidirectional streamline primitives based on motion path or motion direction and maintaining topological relations based on its real or schematic lines and other flow lines. The constructing core for this method is that existing mature theory and application. For example, distance topology catogram construction generation algorithm(Shimizu & Inoue 2009), diffusion-based method for producing density-equalizing maps generation algorithm(Gastner & Newman 2004), genetic algorithms-based equivalent classification algorithm(Armstrong et al. 2003) and graph partition-based multilayers flow mapping generation algorithm(Guo 2009). This paper will not focus

on this kind of algorithms, but using corresponding algorithms to realize construction of thematic maps.

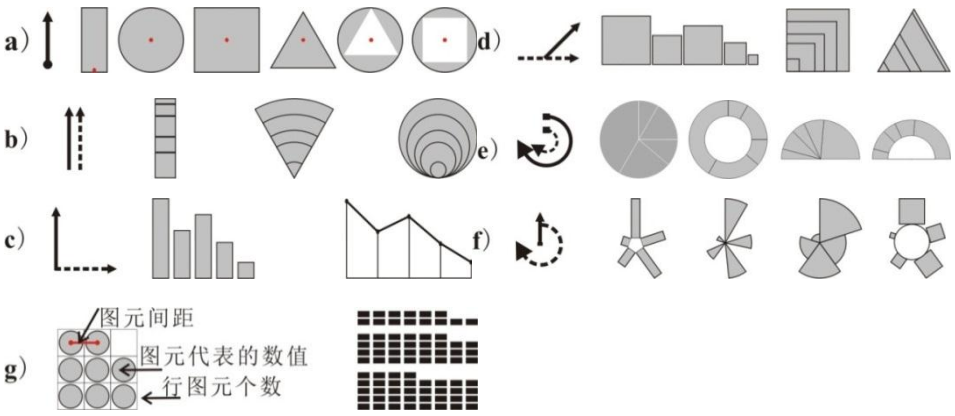


Fig.4 Construction principles in word level

3.3. Combination of syntactic dimension

Symbolic phoneme set S_p based on structural principles of letter level can construct word set S_w with referential meaning. Accordingly, S_w with referential meaning itself and semantic relations between words can construct sentence with O_s., and convey much more complex thematic information. Aiming at different sematic relations between words, there are four categories as following:

1. Synonym combination refers to a kind of combination between words which can convey same or similar thematic elements. A sort of thematic elements can form various expressions through construction principles and retinal variables decoration, then to form aggregation system with all the representation possibilities. In this system, the combination with any two words can form a kind of new express syntax. Different combination with words can stress different thematic contents from different views. Take the resident population at year-end as an example. There are three ways to express quantity. One is that choosing rectangular phoneme to form columnar symbol with linear mode. The other is that choosing polygon phoneme based on topological construction to form classification map. The last one is combination of two above. Allocate columnar symbols to the classification map. And then convey the thematic elements with ratio scale and interval scale. Generally, normalize population with area to form classification map of population density, namely synonym combination.

Two or multiple thematic elements with same or similar meaning content can achieve stress effects through combinations generally. The ways of combinations fall into the following: 1) $O_s.Polar + O_s.Linear$ indicates near-synonym. For example, fan in polar coordinates form to pie diagram strip rectangular linear, and then form to stacked column symbol. Pie chart can highlight structure attribute, columnar symbol can highlight quantative attribute, and these two above can stress thematic elements; 2) Os. Polar + Os. Polar refers to synonym and near-synonym. For example, fan-shaped construct two semi-circles through clockwise direction and counter clock wise direction. The change of shape or color, or different numerical weights can emphasize different elements; 3) Os. Polar+ Os .Topological refers to synonym and near-synonym. For example, fan-shaped through polar coordinates construction express construction of thematic elements, and geometry polygon of thematic elements through topological construction and adopting classified method construct classification maps to indicate contrast of quantitative spatial distribution.

Figure 5a indicates statistical information of basic education in some administrative area. It is a combination of two semi-circles from top to bottom. The top semi-circle indicates quantity and composition of basic education agencies in this administrative area, the bottom semi-circle indicates student enrollment and composition of basic education agencies in this administrative area. We should note that the meanings of synonym combination is mere similar but not same. In the instances above, the number of schools and students has relativity. In general, if there are more schools, the students will be much more. So using semi-circle to contrast, radius ration may effect on cognition.

2. Hyponymy combination includes hierarchical combination and same class combination. The former one is a combination of superordinate and hyponym, it indicates the degree of hyponym which includes in superordinate. For instance, combination express of agriculture and farming indicates the condition of farming in agriculture. The latter one is that hyponym combination indicates superordinate in the same class. For example, combination express of farming animal husbandry refers to agriculture development in the same level. Two kinds of combination uses polar coordinate construction in general. Because quantity unit between words may be not unanimous, coordinate axis of linear construction is unique whether vertical combination or horizontal combination. The ways of combination shows below: $O_s.Polar(\cdots, Polar.startAngle, \cdots) + O_s.Polar(\cdots, Polar.startAngle + s_{pj}.Angle^*p, \cdots)$, and $p \in (0,1)$.

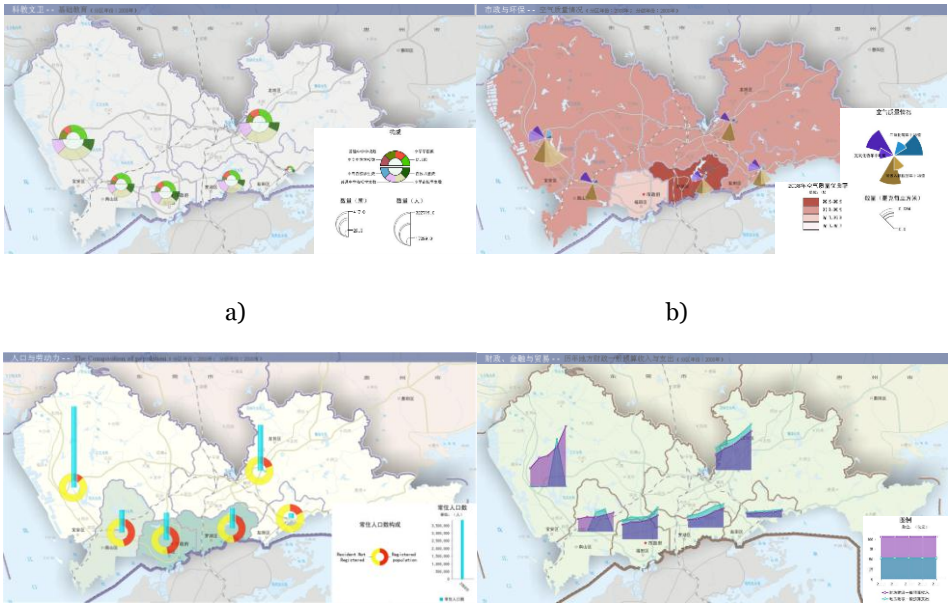
Figure 5b shows atmospheric condition over years in some city, it has three indexes which are sulfur dioxide, nitrogen oxides and particulate matter.

Three indexes as hyponym vertical combination convey the hyponymy relations of superordinate which is atmospheric condition. It includes two combinations. One is that three indexes in one year are generated by polar coordinate construction of fan-shaped morpheme, different sizes of scales indicate different quantative attribute of quantity unit and similar color zone indicate different hyponymy; The anther is that when structuring polar coordinate in different years, changed start angle combined with the last year's and brightness indicates time attributes.

3. Meronymy combination can show part-whole relation of thematic contents or between contents. It includes whole data feature and part of structural feature, reflexes whole quantative feature through linear construction (or fixed construction), and polar coordinate describes feature partial structure. Finally, the two will be generated by fixed construction. Its basic combination formation is: O_s.*Point*(O_s. *Polar*, O_s. *Linear*). In Figure 5, columnar symbol which refers to total population is generated by fixed construction, and ring symbol refers to compositions of total population. In this combination, dimension variable is used in "whole word" to reflect whole numerical of thematic elements, and "part word" adopt color variable to distinct composition feature, and its size does not have numerical meaning.

4. Antonymy combination mainly refers to the contrary relation of objects. It can be achieved prominent effect or counterbalance effect by contrasting. Antonymy relation mainly reflects binary antonymy (such as male and female in sex composition) and relation antonymy (financial revenue and expenditure). Antonymy combination adopts linear construction generally. When conveying binary antonymy , data direction of words and construction should meet at right angles and get bipolar orientation such as pyramid chart. When reflecting relation antonymy, the direction of data should be same, such as linear statistical chart and planar charts. Its basic combination is: O_s.*Point* (O_s.*Linear*, O_s.*Linear*).

In thematic maps, pyramid symbol which reflects population composition of age and sex adopts linear construction to describe different age groups of male population and female population. Its direction of data is opposite with construction and then constructed by fixed construction to manifest contrast for different groups of male and female. In Figure5, combination of different plane domain word symbols reflects financial revenue and expenditure recent years. This combination can notice financial condition easily, purple represents local general budget revenue, green represents general budget expenditure. From the figure, revenue is more than expenditure, and increase every year.



a) b)

c) d)

Figure 5. Combination principles in sentence level

4. Conclusion

The proposed theory provides a new idea for symbol construction and type of innovative. And it has been successfully applied to the Online Dynamic Atlas of Shenzhen(Fei et al., 2011) and Zhejiang Online dynamic thematic map production software in China. Furthermore, online interactive mapping model and web thematic cartography services have theoretical and realistic reference meaning to the automatic cartography for distributed data.

In this mechanism, symbolic construction based on three levels which are letter, word and sentence. And each level can convey dynamically and intelligently thematic essential factors of many dimensions through integration for construction principles of primitives and retinal variables. For further research, automatic recognition and intelligent selection of multiple semantic spatial information, and construction mode deserve more attention. And, recognition effects of construction symbols based on this mechanism and measurement for amount of information are also significant.

References

ARMSTRONG M P, XIAO N, BENNETT D A (2003) Using genetic algorithms to create multicriteria class intervals for choropleth maps [J]. Annals of the Association of American Geographers, 93(3): 595-623

BERTIN J (1983) Semiology of Graphics: Diagrams,Networks,Maps [M]. University of Wisconsin Press; Madison.

BOULOS M N K, WARREN J, GONG J, et al. (2010.)Web GIS in practice VIII: HTML5 and the canvas element for interactive online mapping. International Journal of Health Geographics [J], 9: 1-13

Fei Z, Qingyun D, Zifeng P, et al.(2011) Interact ive Model for Web Thematic Cartography: A Indicator-driven and Task Flow-centered Approach [J]. Acta Geodaetica et Cartographica Sinica, 40(5): 655-661

GASTNER M T, NEWMAN M E J(2004) Diffusion-based method for producing density-equalizing maps [J]. PNAS, 101(20): 7499-7504

GUO D (2009) Flow mapping and multivariate visualization of large spatial interaction data [J]. Visualization and Computer Graphics, IEEE Transactions on, 15(6): 1041-1048

Hongsheng L, Yingjie J,Zhuoyuan Y, et al.(2009) Applying Case-based Reasoning Techniques in Selection of Representation Method of Statistic Map. JOURNAL OF GEO-INFORMATION SCIENCE, 11(6):819-825

IOSIFESCU-ENESCU I, HUGENTOBLER M, HURNI L (2010) Web cartography with open standards - A solution to cartographic challenges of environmental management. Environmental Modelling & Software [J], 25: 988-999

JANKOWSKI P (2009) Towards participatory geographic information systems for community-based environmental decision making. Journal of Environmental Management [J], 90: 1966-1971

JING T, RENTAO H, QINGSHENG G U O (2007) Study on intelligent choice of representation methods in thematic map. Science of Surveying and Mapping [J], 32: 170

Qingyun DU (2004) The Microlinguistic Conceptual Model of Spatial Information. GEOMATICS WORLD ,2(6):5-20

SCHNABEL O, HURNI L (2009) Primitive-based Construction Theory for Diagrams in Thematic Maps. Cartographic Journal [J], 46: 136-145

SHIMIZU E, INOUE R (2009) A new algorithm for distance cartogram construction. International Journal of Geographical Information Science [J], 23: 1453-1470

Wei L, Yongqian Z (1997) The Design and Implementation of a Symbol Base in Thematic Mapping[J]. Journal of Wuhan Technical University of Surveying and Mapping, 22(3): 263-265

WILKINSON L, WILLS G (2005) The grammar of graphics [M]. Springer Verlag

Xiaoning J (1986) THE POSITIONAL SYMBOL SYSTEM OF COMPUTER-ASSISTED THEMATIC MAPPING AND AUTOMATIC DRAWING OF DERIVED SYMBOLS[J]. Acta Geodaetica et Cartographica Sinica, 15(3): 196-206

Yaofeng M (1997) A study on the symbolic construction elements of thematic map[J]. GEOGRAPHICAL RESEARCH, 1997, 16(3):23-31

Yixin H (1993) Determine Map Content Of Thematic Map With Expert System Technology[J]. Journal of Geomatics Science and Technology,10(1):56-61