

Data coding for indigenous language research: attaching local meanings in generating categories and themes

Roselind Wan^{1*}

¹Department of Management and Humanities, Universiti Teknologi PETRONAS, Seri Iskandar, 32610 Perak Malaysia

Abstract. Data analysis in qualitative research is the most complex phase of a study and covers a broad range of approaches with different variation in concepts, assumptions and analytic rules. It involves creating analytic categories reflecting the experiences of participants and highlighting the significance of cultural events happening in the research setting. Accordingly, thematic analysis of data is conducted to examine systematically the content of data to identify recurrent patterns in participants' response and their understanding of their social world. In this paper, I describe the development of categories and themes in a study designed to examine the perceptions on language and identity of Kayan elders in Sarawak, Malaysian Borneo. It demonstrates the analytic rigour in the stages of coding, and offers a rationale for initial coding in the indigenous language. This paper concludes with suggestion for considering indigenous language in coding as it enables the researcher to unpack layers of meanings that are embedded in the cultural context, thereby, enhancing richness of the analysis.

1 Introduction

In qualitative research, the data analysis process is considered the most essential aspect of the study. Analysis of data begins with reading and re-reading through the data corpus to make sense of the data (Hammersley & Atkinson, 1997), identifying analytic categories that reflect on the experiences of participants and henceforth, highlighting the significance of cultural events happening in the research setting. Various scholars consider qualitative coding of data as heuristic and an inductive process (Corbin & Strauss, 2015; Emerson, Fretz & Shaw, 2011; Richards, 2015; Saldana, 2009) and that, the coding process itself, is analysis (Miles & Huberman, 1994). In fact, the goal of qualitative coding is to retain the data until they are fully understood and “to learn from the data, to keep revisiting data extracts until you see and understand patterns and explanations” (Richards, 2015:104).

For an ethnographic study, it requires reading through interview transcripts and field notes, and taking the entire record of the field experience as a complete corpus for analysis. The heuristic nature of data analysis, hence, sees coding as an integration of sequences from observations; “a global reference which encompasses these observations and within which the different data throw light on each other” (Baszanger and Dodier, 2002:11). In which case, the researcher needs to have a certain perspective on observations made in the field so that the interpretations, and the field-notes, are in dialogue with each other (Ricoeur, 1992). This means, the researcher needs to understand the relationship between the how/what and the why of the phenomenon in the research. Nonetheless, qualitative research covers a broad range of approaches to data analysis with different scholars offer variation in concepts, assumptions and analytic rules based on their epistemological and ontological assumptions on research.

In this paper, I describe the development of categories and themes in a study designed to examine the perceptions on language and identity of Kayan elders in Sarawak, Malaysian Borneo. I present a rationale for using Kayan language in the initial coding, suggesting that coding done in participant's language may offer a viable alternative to deeper analysis of the data. Since the goal of ethnography is to “grasp the natives' point of view, his relation to life, to realise his vision of his world” (Malinowski, 1922: 19), coding the data in Kayan gives preference to participants' own words, and thus, retaining the essence of the native's meaning. For this paper, materials for data analysis are selected from the corpus of a larger ethnographic study that consists of data from interview transcripts, participant observation field notes, journals, memos, artefacts, songs, photos, folk tales and speeches. The aim is to demonstrate analytic rigour in qualitative data analysis and offers an alternative method on theme construction that can be used in similar studies. Readings from grounded theories (Corbin & Strauss, 2015; Emerson, Fretz & Shaw, 2011; Saldana, 2009) and ethnographic studies (Baszanger & Dodier, 2002; Brewer, 2000; Hammersley & Atkinson, 1997) informed my understanding and suggestion for theme development process in this analysis.

* Corresponding author: roselind_wan@utp.edu.my

2 Participants in the study

The participants are Kayan elders from a longhouse in the Baram region in Sarawak, Malaysian Borneo. They live in their traditional longhouses located in the rural regions of the Baram and in Belaga. The Kayan is a changing society. In the past, they are known as a communal society that lived entirely in the longhouse. However, in the past two decades or so, due to changes brought about by education and the pursuit of salaried jobs, many have left their longhouse to live in the towns. Although most still live in their longhouses, it is observed that large majority of Kayan in recent times have migrated to live in the towns. A 2010 census estimated the population of Kayan in Sarawak to be around 27,000, less than 0.04 per cent the total population of Malaysia. Kayan culture is generally deeply ingrained in the cultural landscape of Sarawak, and their traditional music and dance given notable attention in contemporary writings. It is unfortunate however, scholarly literature on the Kayan is still under developed. In particular, post-colonial writings on issues pertaining the Kayan language, and the social and cultural changes being experienced by the Kayan do not received sufficient attention from scholars. For this reason, my study hopes to fill the gap in research and contribute not just to our understanding of the Kayan but more importantly, to the much-needed scholarly literature on the Kayan.

The main sources in this study are participants from a Kayan longhouse located in the Baram district. I categorized them into two different groups of participants whom I refer to as elders (see Figure 1). For indigenous communities, elders are often older members of the community; individuals who are recognized as having spiritual and cultural wisdom and possess knowledge of traditional ceremonies, stories, and teachings from centuries past (Mercurieff & Roderick, 2013). Based on Mercurieff and Roderick (2013)'s contention, I consider those above fifty years of age as elders. The first group of elders, that is, Group 1 are those whom I classify as *daha aleng melo uma* [those who stay at home] and the second group of elders (Group 2) are *daha aleng melo ha'oh* [those who stay in the towns]. *Daha aleng melo uma* are elders that live their whole life in the longhouse. Elders in *daha aleng melo ha'oh* group were selected from those who work and live in the towns. Through my inquiries during fieldwork, this second group of elders has regularly returned to the longhouse over the past several years and they maintain close relationship with the Kayan in the longhouse. I chose these two groups as I anticipate that data from both may offer interesting insights on Kayan perception on their language and identity in Sarawak.

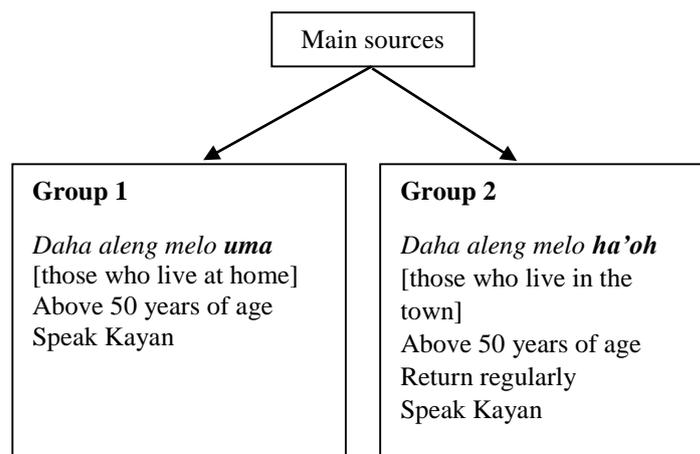


Fig. 1. Participants for the study

3 Understanding meanings in local context

It seems appropriate at this juncture to give a little explanation on the words *uma* and *ha'oh*. According to Fetterman (1998) and Spradley (1979), native words are rich in cultural meanings, thereby giving the English translations for the word *uma* [house] and *ha'oh* [downriver] in coding, may not fully represent their local meanings. For example, *uma* for Kayan is the longhouse but which also means home. In fact, *uma* unlike the English connotation, implies more than just house and home. According to Basso (1988), it is important to understand the environment, and perhaps the meanings and significance of the landscape, and that of the speech acts of the informants so that these could be interpreted in the manner that is fair and near to the meanings that locals attached to it. The elders have lived most of their lives in the longhouse. For them, *uma* suggests a connection to a sense of place, a place with whom the Kayan have an intangible spiritual bond (Basso, 1988), in that it conjures a feeling that is resonance to the mention of *asen* [root, origin], family, the longhouse and home.

Concomitantly, for those who live *ha'oh*, *uma* is often associated with *liveng* - a yearn for the past, for reflection into their memories of growing up in the longhouse. The Kayan in the second group may have a house in the cities, but home to them is always at the *uma* [the longhouse]. The word *ha'oh* in itself means 'downriver' (Southwell, 1990) which refers

to any place downriver from the longhouse. In the distant past when travels were limited due to inaccessibility and having no means of communication between the longhouse and the towns, *ha'oh* means a place far away. The isolation of the longhouse from other communities made any place *ha'oh* as somewhat mysterious and out of reach. In contemporary reference, *ha'oh* is the town or cities where young Kayan *ilo hadui* [look for work] or where they *ilo urip* [look for life]. The significance of these terms is ingrained in the meanings that only members of the community know, although sometimes, and sadly so, these meanings are taken for granted (Basso, 1988). Understanding the meanings of words in their local context brings insightful layers to coding and enhance the qualitative analysis of a study.

4 Data Collection

Specifically, for this paper, data is gleaned from interview transcripts of only Group 1 (Table 1). In the extended analysis of the study, data from Group 1 is supplemented with data from Group 2, and from speeches and field notes taken during a study spanning over four years of intermittent field work. Two elders (*) passed away a few months after the interview. I have decided to retain their interview data as they hold valuable insights that complement the other data sets.

Table 1. Profile of elders in Group 1 – Daha aleng melo uma

| Profile | <i>Usun</i> | <i>Jau</i> | <i>M. Ing*</i> | <i>Mok Ya*</i> | <i>Ulau</i> |
|----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Age | 64 | 75 | 76 | 80 | 64 |
| Gender | F | M | M | F | F |
| Education | Primary School | Primary School | Primary School | - | Primary School |
| Occupation | Farmer | Farmer | Farmer | Farmer | Grocery Shop owner |
| Language | Kayan | Kayan | Kayan | Kayan | Kayan |
| Location | Long Panai |
| Data collection type | Ethnographic interview |

Data collection occurred in two phases (see Figure 2). In Phase One, I conducted ethnographic interviews lasting between 1 – 2 hours each for all the elders. As the elders are known personally to me, the nature of the interview is informal and although it is different from conversation, for ethnographic interviewing, the two would typically merge into one (Fetterman, 1998). Elsewhere in my thesis, I refer to this amalgamate as *tengaran*, a Kayan term for talk and conversation which can be both formal and informal. In Phase Two, data collection was through participants' observations and speeches. This includes field note observations, memos and recorded speeches that supplement data collected through interviews. The data corpus thus offers me insightful data on the social world of the community.

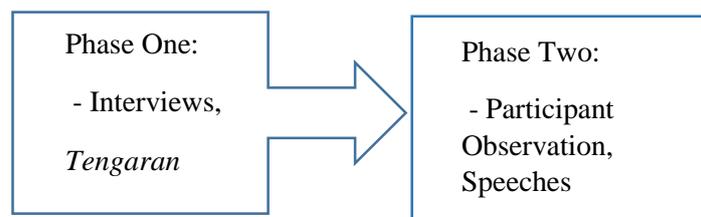


Fig. 2. Phases of data collection

5 Coding in Kayan language

Smith, Chen and Liu (2008) say it is helpful to develop a coding framework in the same language for qualitative interviews that are conducted in the local language. This is because original words, phrases and concepts are securely embedded in

the local context and the risk of misinterpretation and loss of participants' intended meaning can be minimised. Proper management of the data ensures that the credibility, transferability, dependability and confirmability of data are not compromised. Identifying an initial coding frame requires skill in pinpointing recurring themes and concepts and developing meaningful labels for the data. To do this, I need to become familiar with the data through listening attentively to the audio recordings. From here, I was able to gather some general sense of emerging categories that indicate possible relevance to the research questions. The interview data was transcribed and then translated into English. Prior to analysis, I read each transcripts a couple of times and wrote analytical memo in Trello, an organizing app to help me have a visual overview of main ideas and emerging categories. I then started coding the English transcripts. Text segments (or meaningful units) were compared and contrasted and assigned inductive codes.

However, on hindsight, I discovered it would be much better to code in Kayan because some nuances of elders' meanings are not fully captured in the English version of the transcripts. For instance, as I highlighted in the earlier section, there are words and expressions from elders that given an English translation would lose some of their dynamics. Complexity also arise in translation, in particular when no equivalent word exists in the target language including the influence of the grammatical style on the analysis (Twinn, 1998). It is thus necessary for me to avoid problems of interpretation and to ensure accurate meaning of data. Also, when the coding is made in Kayan, the meaning will not be lost, the essence of the participants' voices and their cultural nuances are still being retained.

In my study, there are many such examples of words and phrases in the indigenous language that are embedded with local meanings, for instance, the phrases *dahok mahen* and *dahok murah* in the example below. Given a loose English translation, *dahok mahen* and *dahok murah* means 'expensive and cheap language'. This simple translation does not quite capture the layers of meaning that the Kayan word entails. The nuances of *dahok mahen* and *dahok murah* necessitates understanding the concept, and in what context the word is being used, as in Figure 3:

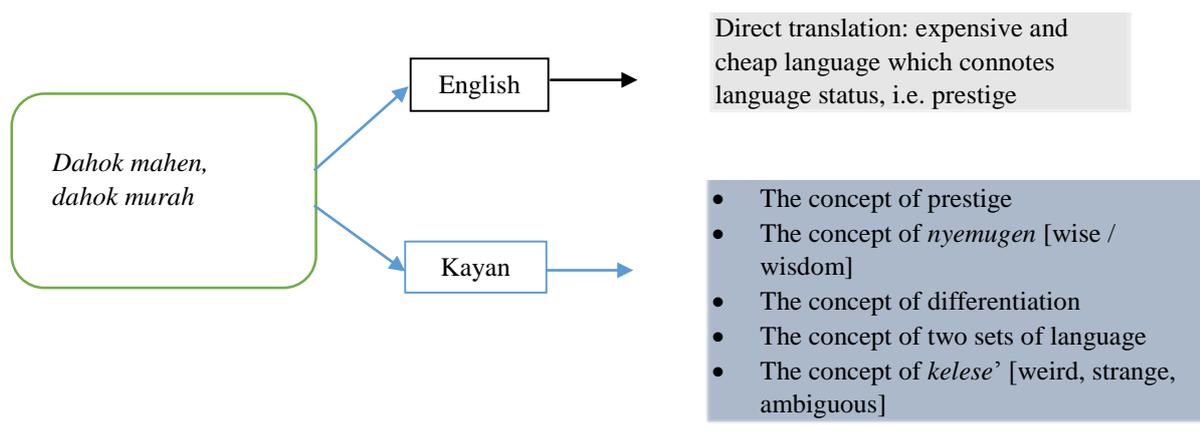


Fig. 3. The concept of *dahok mahen*, *dahok murah*

Using the Kayan transcripts for open coding allows for layers of meanings which an English translation would not have been able to capture. Here, given a coding in Kayan, the concept of *dahok mahen* and *dahok murah* encapsulates several layers of meaning. More significantly, it assumes the Kayan language to be of two types; old Kayan and 'contemporary' Kayan. That is, the old Kayan language is seen as a *dahok mahen* [prestige language], and often associated with the *dahok menuna* [language of the ancestors]. The *dahok mahen* is therefore considered pure and untainted, free from corruption. Further, their oral tradition *tekna'*, is sung in the old Kayan, where those who can sing the *tekna'* are often viewed as people who are knowledgeable and possess the wisdom of the ancestors. However, the contemporary Kayan language is *dahok murah* [cheap language] and *kelese'* [weird, strange and ambiguous]. The elders refer to it as *dahun nyam kere nih* [the language of the young people]. Consequently, there is the perception that those who do not know how to sing and understand the *tekna'* as quite illiterate in the *ju* Kayan [the Kayan way]. From this example, we can see how coding in Kayan offers voluminous potentials for meanings to be uncovered where *dahok mahen* and *dahok murah* illuminate the notion that 'old' Kayan is a language of prestige and tradition, whilst the contemporary Kayan as one which is contaminated and polluted due to language mixing.

6 Developing categories and themes

The text segment in Kayan transcript below (Table 2) is taken from an interview with one of the elders to illustrate initial coding process in the indigenous language. Making the analysis in the words of the elders allows for exploration of meanings as experienced by the elders themselves, therefore, a richer understanding on their cultural meanings can be achieved (Spradley, 1979). As can be seen, even from this short segment of data, it is possible to come up with a variety

of codes from which relevant categories and themes will emerge. Descriptive and in vivo codes are applied to the text segment (Saldana, 2009). There are similar recurrent themes emerging from the data which I grouped into three categories: *meng melak*, *levah*, *asen* [do not abandon; the language will disappear or die; the language is our root or identity]. These categories would later be refined with other categories from other text segments to form higher level categories:

Table 2. Initial coding in Kayan

| <p>Talk data 5 Source Mok Ya @ Aug 2014</p> | <p>Codes (descriptive and in vivo coding)</p> |
|---|---|
| <p><i>bei, marung</i> ¹<i>dahun kayan nih, sayu lelan kah kurin, kenep ta re, nureh yah...</i> ²<i>Nusi tam deng melak na. ita em jam ikam nyam nyam nih, ita anih tei pah na'a, daha anak kam anan usi nah ita jam. Pah si ya, ikam na na na ure' pah</i> ³<i>iha seng te'.</i> <i>Nutih.</i> ⁴<i>Nei pa ji nganak pa dahun uma beh, nei pah ji nganak pa dahun uma beh, dahun inggeris dahun melayu daha tih pah,</i> ⁵<i>tam lak2 nah dahun tam jam na nipa. (avin na nun tam melak?) bei, avin</i> ⁶<i>ha kajo dahok atih lo kereh, nuno,</i> ⁷<i>daha duyah dahun tam yah, bi kurin ta anih sayu lelan dahun tam nih itam na pa'en. Kurin ta nih. Usi dahun tam nih seng tei</i> ⁸<i>levah kurin ta nih.</i> ⁹<i>Menangen tam na kah kurin ta nih, bi ha dap kerei nih pa, na dahun ha beh lelan daha men anak daha.</i> ¹⁰<i>Anan pah ji tusah nih pah. Menangen kelo uk te dahok kayan anan</i> ¹¹<i>ja ha sayu kereh.</i> ¹²<i>Ja ha em tei pah,</i> ¹³<i>tei melak.</i> ¹⁴<i>Bara asen tam atih ka sayu na.</i> ¹⁵<i>tei ha melak</i> ¹⁶<i>usi nah te dama asen tam la'an ujung</i> ¹⁷<i>asen daha tua nah. Nure' pelemana kui reh. usi ta mun melak na urei ha em tei levah.</i> ¹⁸<i>avin tada tam kayan atih pah, tei ha levah, ujung itung</i> ¹⁹<i>dahun uma beh tua nah</i> ²⁰<i>usi gunah nutih,</i> ²¹<i>bi te lan pu'un tam kayan anih,</i> ²²<i>meng na na levah,</i> ²³<i>na pa'en, nutih aleng pelemana kui reh</i></p> <p>[Regarding our language, it is really good, I think... it is not good for us to <i>melak</i> it. I do not know about you, young people, but when we are gone, I won't know what will happen to the language. May be it will be gone, finished. Unless you speak it. It is like that. It seems that, parents speak other languages (English or Malay) to their children, we have <i>melak</i> our language. We do not speak it. <i>Why do you think that we do not speak it?</i> Well, is it because the other languages are <i>kajo</i>? why is it that they do not want to speak our language, but to me, our language is <i>sayu lelan</i>, that we should speak it always. I would hope that our language will not <i>levah</i>. We should always speak it, speak it <i>menangen</i> to our children, grandchildren. That is the problem. We should <i>menangen</i> speak it. But it's different now. People just speak other languages to their children. This is the problem. You must <i>menangen</i> speak the Kayan language so that it is <i>sayu</i>. So that it will not <i>pah</i>, <i>levah</i>. Otherwise we will just speak other people's languages. It is <i>sayu</i> to show our <i>asen</i>. This is what I think about it. We should not <i>melak</i> it so that it will not <i>levah</i>. Because it is <i>sayu</i> to show our <i>asen</i>, a <i>tada</i> of our identity. If it is <i>levah</i>, then our <i>asen</i> is no longer as many, it would only be other people <i>asen</i>. This is what I think. Do not <i>melak</i> it so that it will not <i>levah</i>. Because it is our <i>tada</i> as Kayan. What is the point, if we end up just speaking other <i>uma ha beh</i> languages because we have our <i>pu'un</i> as Kayan? Do not let it <i>levah</i>, speak it always. This is what I think].</p> | <p>Dahun Kayan sayu lelan¹ Usi deng melak² Ikam nyam na na ure' seng te³ Tei pah na'a⁴ Tam lak avin pekunak dahun ha beh⁵ Kajo sik daho ha beh⁶ Daha duyah⁷ Levah⁸ Menangen tam na⁹ Anan pah ji tusah nih pah¹⁰ Ja ha sayu¹¹ Em tei pah¹² Tei melak¹³ Bara asen¹⁴ En tam lek dahun tam avin na dahun ha beh¹⁵ Usi nah dama asen tam¹⁶ Asen daha tua¹⁷ Tada tam¹⁸ Tei ha melak, usi te dama asen tam la'a¹⁹ Usi gunah nutih²⁰ Te lan pu'un tam Kayan anih²¹ Meng na na levah²² Menangen na dahun tam²³</p> <p><u>Kayan-English</u> Melak: abandon, leave Kajo: beautiful, good Sayu lelan: very good Levah: disappears, gone Menangen: purposeful, persevere Pah: finish, gone Sayu: good Asen: origin, heritage Pu'un: origin, root Tada: symbol Uma ha beh: other people's</p> |

Dealing with the data and coming up with themes is not a straight forward process. In the earlier stages of my data analysis, interesting patterns emerged of possible loose categories (e.g. Christmas, Education, Intergenerational transfer, Land, Modernisation, Oral tradition, Social structure, Longhouse, Complexities, etc.). These loose categories form the basis for development of later categories and are helpful as they offer early analytic insights of what elders perceived as meaningful in relation to Kayan language and identity. The iterative nature of analysis (Fetterman, 1998) as such, makes data analysis in qualitative research both challenging and interesting. It demands that the researcher makes choices “between logical and enticing paths, between valid and invalid but fascinating data, and between genuine patterns of behaviour and series of apparently similar but distinct reactions” (pp 92).

Hence, developing categories from the codes previously identified was a time-consuming process. The categories of the data need to be clarified and developed in relation to one another, including specifying the links between the various concepts and indicators derived from the coding. In the text segment example presented above, 23 codes were identified. These codes need to be clarified with the other codes from the other data sets, which altogether for Group 1, there are about 300 codes. These were later categorized into 58 higher order themes. From the Kayan codes, categories were refined and regrouped into five categories and further refined into themes (See Figure 4):

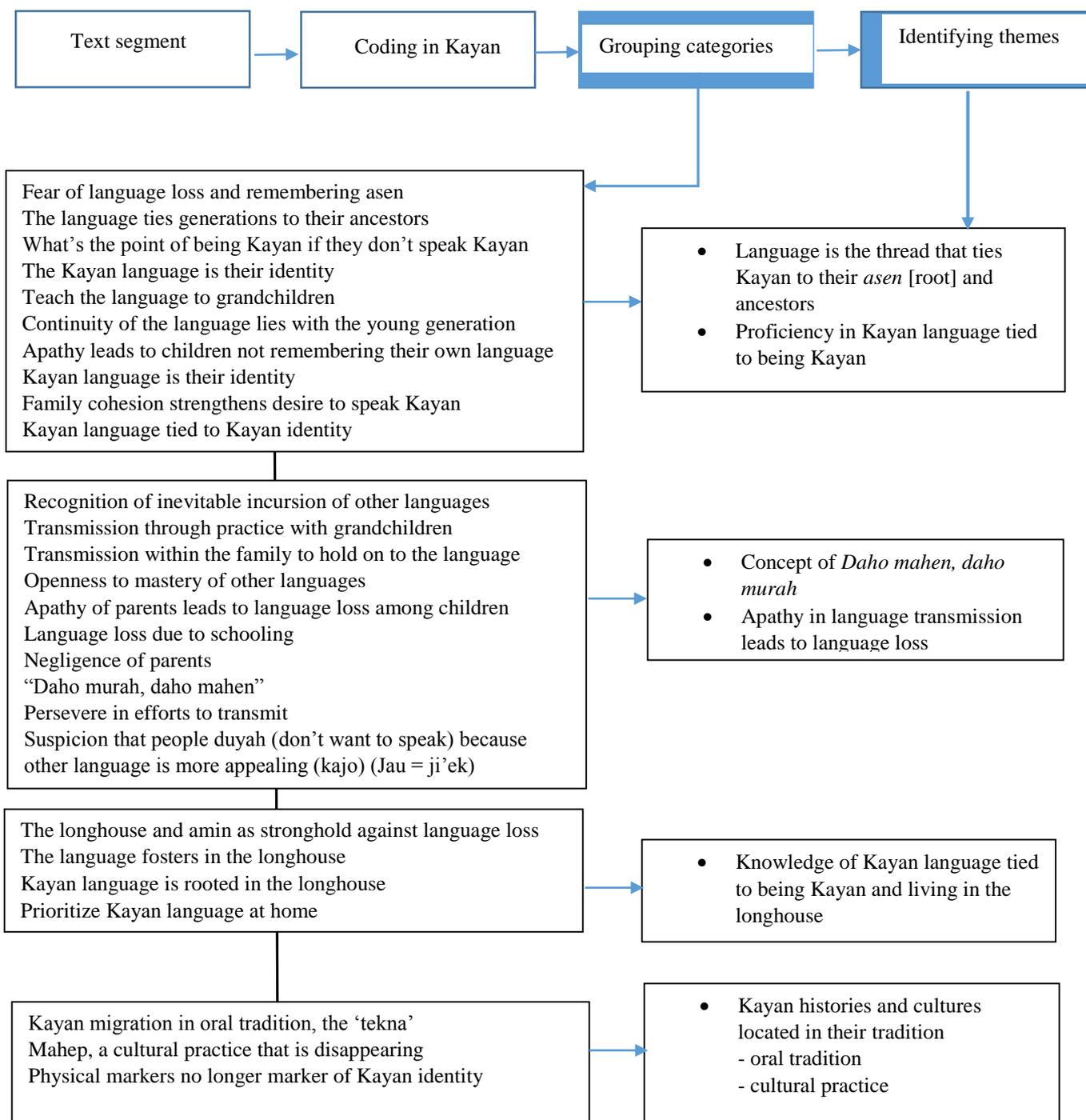


Fig. 4. Development of codes into categories and themes

7 Conclusion

Qualitative coding in the indigenous language open avenues of inquiry through identifying codes using the elders' own words without translating the text into English. Retaining the elders' words and coding in the original language enable the researcher to unpack layers of meanings that are embedded in the cultural context, thereby, enhancing richness of the analysis. It is possible that translating and coding in a translated form could have given a different interpretation to the

text. Qualitative analysis is also enhanced when the researcher considers the real-world context of participants in the data analysis process. This paper has offered an alternative way of initiating the initial coding of a qualitative study in the Kayan language. It highlights that coding done in the original language of the elders could offer richness to the interpretation of the data while at the same time, making sure that the analysis still agree to the context from which data were collected.

8 References

- Baszanger, I., & Dodier, N. (2002). Ethnography. Relating the Part to the Whole. In *Qualitative Research: Theory, Method and Practice*, pp 8-23. D. Silverman (editor). London: SAGE Publications Ltd.
- Basso, K. H. (1988). "Speaking with Names": Language and Landscape among the Western Apache. *Cultural Anthropology*, 3(2), pp 99-130. Blackwell Publishing
- Brewer, J. D. (2000). Ethnography. In *Understanding Social Research Series*. A. Bryman (editor). Buckingham, UK: Open University Press
- Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th edition). Thousand Oaks, CA: Sage Publications Ltd
- Fetterman, D. M. (1998). Ethnography. Step by Step (second edition). *Applied Social Research Methods Series*, 17. Sage Publications
- Hammersley, M. & Atkinson, P. (1997). *Ethnography. Principles in Practice* (2nd edition). New York, NY: Routledge
- Illarion L. Mercurieff & Libby Roderick (2013). *Stop talking. Indigenous ways of teaching and learning and difficult dialogues in higher education*. University of Alaska Anchorage.
- Malinowski, B. (2002). *Argonauts of the Western Pacific. An account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. London: Routledge
- Miles, M. B. & Huberman, A.M. (1994). *Qualitative Data Analysis* (2nd edition). Thousand Oaks, CA: Sage Publications.
- Ricoeur, P. (1991) *From Text to Action: Essays in Hermeneutics, Vol. 2*. Evanston: Northwestern University Press.
- Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage Publications Ltd.
- Smith, H. J., Chen, J., & Liu, X. (2008). Language and rigour in qualitative research: Problems and principles in analysing data collected in Mandarin. *BMC Medical Research Methodology*, 8, 44. Downloaded March 27, from <http://doi.org/10.1186/1471-2288-8-44>
- Spradley, J. P. (1979). *The Ethnographic Interview*. New York: Holt, Rinehart and Winston, Inc.
- Southwell, C. H. (1990). *Kayan-English Dictionary*. Kuching, Sarawak: Sarawak Literary Society
- Twinn S. (1998). An analysis of the effectiveness of focus groups as a method of qualitative data collection with Chinese populations in nursing research. *Journal of Advanced Nursing*, 8, 654–661. doi: 10.1046/j.1365-2648.1998.00708.x