Open Access

Main Concept Analysis for Acquired Deficits of Spoken Narratives: Preliminary Data on Inter-rater Agreement and Potential Application to the Korean-Speaking Population

Anthony Pak-Hin Kong

Department of Communication Sciences and Disorders, University of Central Florida, Orlando, FL, United States

This study aims to investigate the inter-rater agreement of the Main Concept Analysis (MCA), a proposition-based system for analyzing the presence, accuracy, completeness, and efficiency of content in spoken narratives of speakers with aphasia. Twenty-one MCA assessments were administered to thirteen participants recruited from an intensive aphasia treatment program. Six MCA indices were applied to the language samples, which were cross-checked to determine discrepancies of results across raters. The present results were consistent with similar studies in the literature, thus indicating that the Main Concept Analysis is a reliable assessment battery. Given the simple, quick, but objective procedures for language quantification, it is argued that the Main Concept Analysis can easily be adopted to the Korean-speaking population for clinical analysis of discourse.

Keywords: Main Concept Analysis, Oral discourse, Scoring agreement, Aphasia

INTRODUCTION

Aphasia is a type of acquired language disorder in adults as a result of a cerebrovascular accident (i.e., stroke) or traumatic brain injury (TBI). People with aphasia (PWAs) or adults with TBI demonstrate impairments in language production, such as reduced content and lexical diversity, syntactic complexity, impaired cohesion and coherence [1]. Discourse level analysis of oral production by PWAs is informative of their micro- and macro-linguistic analysis. However, detailed and comprehensive analysis of spoken narratives is clinically infeasible due to the extensive amount of time involved in sample elicitation and especially the subsequent analysis [2]. In addition, most existing standardized language batteries of aphasia or TBI do not contain detailed and systematic evaluation of oral narratives. As a result, clinically oriented tools that were designed to assist clinicians in quantifying disordered spoken narratives became more popular. Specifically, the results of these clinical tools can assist healthcare professionals in making an initial evaluation of a speaker's linguistic abilities, which may then lead to more sophisticated and in-depth assessment of aphasia.

Various clinically feasible approaches of discourse analyses, each with unique methodologies and, therefore, strength and weaknesses, have been proposed (see detailed

review in Kong [1]). One of them is the measurement of main concepts [3], which is a propositional-based analytic system that focuses on the quantification of presence, accuracy, and completeness of essential information in spoken narratives in PWAs. According to Nicholas and Brookshire [3], a main concept should contain only one main verb and provide an outline of the gist depicted in a picture, or an outline of the essential steps in a procedure. Following this particular analytic approach, the *Main Concept Analysis* (MCA) [4] has recently been developed and published to allow a more comprehensive and multilevel coding system of spoken output by individuals with aphasia [5, 6], dementia [7], and TBI [8] in everyday clinical settings. In particular, the modification included development of four sets of sequential black-and-white line drawings and proposal of new quantification measures.

There are four sets of sequential pictures in the MCA that are utilized for elicitation of language samples. Each picture set, composed of four detached single black-and-white line drawings with a dimension of 15 cm by 21 cm, contains a theme or storyline (plot of story) that is made up of unique lexical items. To be specific, while the first two picture sets contain only one character, the remaining two sets contain three characters and depict more target main concepts than the first two sets. Figure 1 below shows an example of one of the four picture sets (Set 1) that depicts an old lady cooking in a kitchen. Set 2 depicts a theme of a man waking up late for work and wearing a pair of socks that are not matched in color. Set 3 portrays a plot of a girl and her mother getting ice-cream. Finally, set 4 shows a boy helping an old man carrying whose grocery bag filled with oranges is broken. While there is only one character in picture sets 1 and 2, each of the other two sets contains three characters. These pictures were designed to elicit discourse with lexicons, chosen to be appropriate to speakers of both the eastern and western culture, have been controlled to avoid overlap of use in more than one picture set. According to the author, a speaker should easily identify and convey a visual image represented in these stimuli.

The MCA utilizes six indices to quantify the presence, degree of accuracy and completeness, as well as efficiency of one's spoken narratives. The first four were directly adopted from Nicholas and Brookshire [3]:

(1) Number of Accurate and Complete concepts (AC): All the pieces of essential information in a target main concept, including the characters, their actions, and the objects depicted in the pictures, must be correctly mentioned by a speaker. All the lexical items used to refer to the essential
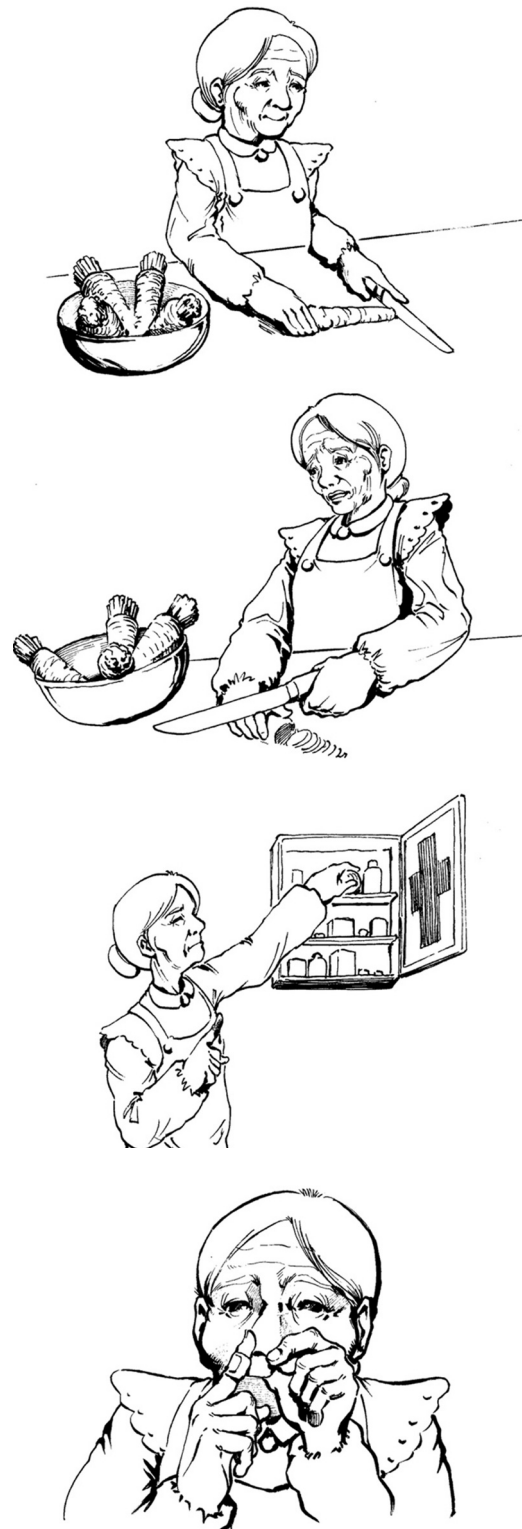

Figure 1. An example of the sequential pictures in *Main Concept Analysis* (MCA) - Set 1 (Cooking in a kitchen). This picture set depicts an old lady preparing a meal. The woman is cutting up some carrots in a kitchen. After cutting her finger by accident, she takes out a Band-Aid from the first-aid kit and covers her wound.

information must be correct.

(2) Number of Accurate but Incomplete concepts (AI): target essential information is only partially provided in a main concept, with the missing of one or more pieces of essential information. All the given essential information must be correct; i.e., if one or more pieces of the target essential information are omitted, a main concept will be scored as AI only when all the remaining content is accurately described.

(3) Number of Inaccurate concepts (IN): This is counted when one or more pieces of essential information in a main concept given are incorrect. Inaccurate information can be given in the form of a semantic paraphasia, phonemic paraphasia, unrelated paraphasia, neologism, or jargon. The fact that a speaker provides incomplete essential information would not affect the scoring because this measure only focuses on the accuracy of the description.

(4) Number of Absent concepts (AB): This is counted when a speaker fails to mention a particular main concept in the oral description or when none of the target essential information within a given main concept is present.

(5) Main concept score (MC score): This is a composite score summarizing the first four MCA indices and is computed by the formula "$3 \times AC + 2 \times AI + 1 \times IN$". The scoring method is based on the considerations of the following three areas of discourse production: the presence of essential information (independent to the degree of correctness) in a description, the accuracy in providing essential information, and the completeness of essential information given. For each main concept, one point is credited towards the final MC score if each of these discourse skills is evident in the language sample.

(6) Number of Accurate and Complete concepts per minute (AC/min): This is a newly added efficiency measure modified from the Index of Communication Efficiency in the Cantonese Linguistic Communication Measure (CLCM) [9]. The total duration of spoken production is first converted into minutes. The index is then computed by dividing the AC by the total duration.

With reference to the example of sequential pictures in Figure 1, Appendix A shows the five target main concepts of Set 1 in the English MCA, along with the acceptable alternative lexical items based on normative data reported in [7]. There are six, nine, and six target main concepts in MCA picture sets 2 to 4, respectively. The MCA has been reported to distinguish PWAs from unimpaired speakers across various language ad-

aptations, such as English [7], Cantonese Chinese [5], as well as Mandarin Chinese for Taiwan [10]. It has also been translated into Irish-English [11] and Putonghua for Mainland China [8]. The MCA is now being expanded and validated into other language versions, including Spanish, Japanese, Persian, and Brazilian Portuguese, through collection of geographically-specific normative as well as pathological (clinical) data.

Measurement and scoring of PWAs' performance at the single word level (such as confrontation or divergent naming of nouns and verbs) or sentence level (such as a sentence construction task) tends to be more straightforward and can generally yield a higher degree of across-raters and within-raters consistency. Agreement on post-sentential analyses, on the other hand, are likely to be more challenging due to the discrepancy of preparing a discourse transcript and involvement of raters' personal judgment in subsequent quantifications. Oral discourse production commonly involves the interaction of multiple linguistic levels of language, including phonology, semantics, morphology, syntax, and pragmatics. The mutual influence of these various layers of performance can be identified with refined discourse analysis [12]. According to Kong [1], "the process of translating spoken data to an orthographic medium involves the making of multiple decisions along the way"… and … "the degree of how many decisions are made by an examiner in preparing and processing a transcript and how reliable these decisions are will depend on factors such as the type of the discourse tasks, experience of preparing transcription and conducting text-based discourse analyses on the part of the raters, and the nature of the actual analyses (p. 66)". Although the analytic procedures of the MCA were claimed to be straightforward and clinically-friendly [5,7,10], inter-rater agreement of transcript coding (and subsequent MCA scoring) seemed to be affected by factors such as examiners' familiarity of narrative analysis or, simply, the test itself.

OBJECTIVE

Rating and quantification of discourse production often involves some degree of subjectivity on the part of clinicians (or test administrators). Given the lack of studies in the literature that examine issues around examiners' agreement in conducting discourse analysis, the aim of this study is to investigate the inter-rater agreement of the MCA. Specifically, we were interested to examine if and how the severity of aphasia

as reflected by the Aphasia Quotient of the Western Aphasia Battery (WAB) [13] might contribute to inter-rater discrepancy of testing involving the MCA.

METHODS

Twenty-one MCA assessments were administered to thirteen chronic participants with aphasia (eleven as a result of cerebrovascular accident and two after TBI) from the University of Central Florida (UCF) Aphasia House intensive program [14]. Table 1 displays the background information of our participants. Aphasia Quotients (AQ) were obtained from the WAB [13] that corresponded with each MCA evaluation to provide insight into the severity of aphasia. Four sets of sequential pictures [4] were used by graduate student clinicians at the UCF Aphasia House to elicit language samples, which were recorded and subsequently transcribed by the corresponding student clinicians. The transcribed language samples were analyzed to determine the accuracy and completeness of main concepts that directly relate to the picture sets with reference to the above-mentioned MCA indices. Note that prior to interviewing the participants, each student clinician received training on the specific sample elicitation methods, probing techniques, and index calculation of MCA. The orthographic

transcripts were independently re-analyzed and crosschecked by two independent raters to identify discrepancies of transcription and index computation. Percentage of across-rater disagreement was determined, along with qualitative summary of factors that contributed to the discrepancies.

To determine the potential relationship between MCA scoring disagreement and the participants' severity of aphasia, a Pearson Product Moment Correlation Coefficient between the discrepancy percentage and PWAs' Aphasia Quotient (AQ) of the WAB was calculated.

RESULTS

Out of the total 546 main concepts scored, the across-rater disagreement was 28%, representing discrepancy errors. The top three patterns of discrepancy were, in descending order: (a) IN mis-rated as AI, (b) AI mis-rated as AC, and (c) AB mis-rated as IN (see Table 2 for details displaying the percentage disagreement of MCA index assignment). A chi-square test of independence was performed to examine the relation between picture set and accuracy of rating main concepts. The relation between these variables was insignificant, X^2 (3, N = 546) = 4.274, p = 0.233. There did not seem to be any effects of picture set on the MCA scoring. Disagreements of inter-rater assignment of MCA indices were found across transcripts elicited from all the four picture sets: 11% of overall discrepancies for both sets 1 and 2, and a slightly lower discrepancy of 8% for sets both 3 and 4. The results of Pearson correlation analysis revealed a lack of statistically significant relationship between the severity of aphasia (AQ) and the percentage of discrepancies per assessment (coefficient of 0.277, p > 0.05).

Table 1. Demographic information of participants with aphasia in this study

Participants	Age	Gender	Etiology	Education background	MC Score
1. DelAn	22	F	TBI	Nursing School	43
2. JohSu	52	F	CVA	Master's	26
3. TurRa	69	M	TBI	Juris Doctor degree	66
4. ParWi*	69	M	CVA	Juris Doctor degree	1/4
5. KelHo	74	M	CVA	Doctor of Medicine in Dentistry	48
6. TwiEm*	67	F	CVA	Some high school	20/33
7. ZlaJe*	82	M	CVA	Some high school	40/31
8. ColWa*	23	M	CVA	Some College	49/55
9. BarMa*	68	F	CVA	Doctoral	21/19
10. TelKe*	62	M	CVA	Some College	21/20
11. FerRe*	51	M	CVA	High School	2/0
12. BloRu*	49	M	CVA	Bachelor's	35/40
13. BreFr	63	F	CVA	High School	44

F, female; male, male; TBI, traumatic brain injury; CVA, cerebrovascular accident; MC Score, Main concept score (with a highest achievable score of 78).
*Participants who were retested of the *Main Concept Analysis* (MCA) within two weeks after the initial evaluation.

Table 2. Percentage distribution of across-rater disagreement in assigning MCA indices

Discrepancy type	Number of discrepancies out of 152 main concepts	Total (%)
IN → AI	84	55.26
AI → AC	15	9.87
AB → IN	12	7.89
IN → AB	9	5.92
AB → AI	9	5.92
Others	23	15.13
Total	152	100

AC, Accurate and Complete concepts; AI, Accurate but Incomplete concepts; IN, Inaccurate concepts; AB, Absent concepts.

DISCUSSION & CONCLUSION

In this paper, we have provided a review of the development and validation of the MCA, a proposition-based assessment tool for quantifying deficits of spoken narratives secondary to acquired neurogenic disorders. The inter-rater agreement of measuring content production using the MCA was examined, with results that are consistent with the point-to-point inter-rater reliability of the Cantonese MCA [5]. In reviewing the characteristics of discrepancies that were found between different raters, two major causes of disagreement were identified: (1) examiners failed to give proper credits to main concepts composed of acceptable lexical alternatives (listed or not listed in the scoring manual), and (2) correctly used pronouns for nouns in a main concept (previously uttered in a language sample) were sometime not properly scored as an accurate or complete piece of essential information. Extending the conclusion of Oelschlaeger and Thorne [15] who suggested that low scoring agreement of aphasic output was mainly attributed to inadequacy of scoring rules (rather than competence of examiners), it is argued that users' knowledge about the scoring procedures and scoring criteria is crucial to reliable scoring of MCA. A lack of clear understanding about the definitions (and scoring principles) for indices IN and AI given in the MCA seemed to have contributed to a majority of the across-rater disagreements. Following the argument in Richardson and Dalton [16], it was hypothesized that severity of aphasia (AQ of WAB) would affect the occurrence of discrepancies, because speakers with a lower AQ would tend to produce less complete and accurate spoken narratives that could potentially contribute to more inconsistent scoring. However, the current results failed to confirm this premise. An alternative hypothesis supporting the present findings could be output of speakers with non-fluent aphasia, who are characterized by less syntactically complex output and overall less quantity of output, may be mis-rated more frequently. Additional data are being collected and analyzed to further investigate factors, such as speakers' age, education, etiology of language deficits (e.g., stroke vs. TBI vs. dementia), genre type, or raters' experience on and knowledge of discourse analysis, affecting the inter-rater and intra-rater of spoken narratives in healthy speakers and individuals with acquired language impairments. It is anticipated that the final findings of this study will be of interest to clinicians who perform clinical assessment and management of oral discourse among PWA. With reference to ASHA's principle of Evidence-Based Practice [17],

we also hope that these findings will facilitate clinical utilization of the MCA as well as other comparable approaches of discourse analysis.

In the past decade, there has been increasing study of language and communication disorders of elderly in the Korean-speaking population [18]. A recent review article on speech and language services for individuals with neurogenic communication disorders in South Korea [19] indicated that standardized language assessment tools for Korean-speaking PWA were limited to the Korean adoptions of the *Boston Naming Test* [20,21], *Western Aphasia Battery* [22,23], and *Minnesota Test for Differential Diagnosis of Aphasia* [24]. An aphasia screener, namely *Screening Test for Aphasia & Neurologic-Communication Disorders* [25], has also been published in 2009. As excepted, the above-mentioned Korean batteries were limited in their capacity to conduct a comprehensive analysis of spoken discourse. It is, therefore, proposed that the MCA may become a potential addition to the pool of clinical resources once it is adequately translated, adopted, and validated for used in the Korean-speaking population. The MCA can also act as a supplement to existing discourse analytic systems adapted for Korean, such as the spontaneous speech analysis scales [26], information analysis of story retelling [27], or quantification of Correct Information Units [28]. The advantages of the MCA pictures being designed to be culturally universal and the scoring procedures being linguistically independent across language should be highlighted again here. In other words, given the simple, quick, but objective procedures for language quantification, it is argued that the MCA can easily be adopted to the Korean-speaking population for clinical analysis of discourse. An example of Korean translations for the main concepts depicted in MCA picture set 1 is illustrated in Appendix B. Specifically, these main concepts and the corresponding list of acceptable vocabulary was determined based on the response from ten native speakers of Korean residing in South Korea, including five SLP and five laymen without professional or specialized knowledge in the subject of speech therapy.

ACKNOWLEDGEMENTS

guage-Hearing Association (ASHA) Convention held in Los Angeles, USA. The author would like to acknowledge Dr. Janet Whiteside, former Director of the Aphasia House at University of Central Florida (UCF), for her help in data collection. Special thanks to Dr. Ikjae Im, Research Fellow at UCF Communication Disorders Clinic, who assisted in the translation of MCA into Korean. Data processing and statistical analysis was conducted by Ana Torres and Stephanie Eaton, UCF student researchers and clinicians.

REFERENCES

1. Kong APH. Analysis of neurogenic disordered discourse production: From theory to practice. New York, NY: Routledge Psychology Press; 2016.

2. Armstrong L, Brady M, Mackenzie C, Norrie J. Transcription-less analysis of aphasic discourse: a clinician's dream or a possibility? Aphasiology. 2007;21:355-374.

3. Nicholas LE, Brookshire RH. Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. Journal of Speech and Hearing Research. 1995;38:145-156.

4. Kong APH. The Main Concept Analysis (MCA) for oral discourse production. Hong Kong: The Commercial Press (H.K.) Limited; 2016.

5. Kong APH. The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: a preliminary report. Journal of Communication Disorders. 2009; 42:442-464.

6. Kong APH. The main concept analysis in Cantonese aphasic oral discourse: external validation and monitoring chronic aphasia. Journal of Speech Language and Hearing Research. 2011;54:148-159.

7. Kong APH, Whiteside J, Bargmann P. The Main Concept Analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. Logopedics Phoniatrics Vocology. 2016;41(3):129-141.

8. Gao G, Kong A, Lau K. Production of main concepts by Mandarin-speakers with traumatic brain injury in China: A pilot study. Frontiers in Psychology. 2016.

9. Kong APH, Law SP. A Cantonese linguistic communication measure for evaluating aphasic narrative production: normative and preliminary aphasic data. Journal of Multilingual Communication Disorders. 2004;2(2):124-146.

10. Kong APH, Yeh CC. A Taiwanese Mandarin Main Concept Analysis (TM-MCA) for Quantification of Aphasic Oral Discourse. International Journal of Language & Communication Disorders. 2015;50(5):580-592.

11. Kong APH, Ross A, Pettigrew C. A Main-Concept Analysis for aphasic discourse in Irish-English speakers: adaptation and pre-liminary report. Journal of Clinical Speech and Language Studies. 2012;19:19-43.

12. Prins R, Bastiaanse R. Review: analysing the spontaneous speech of aphasic speakers. Aphasiology. 2004;18(12):1075-1091.

13. Kertesz A. Western Aphasia Battery. New York: Grune and Stratton; 1982.

14. Whiteside J, Kong APH. Participant outcomes on language, discourse & cognitive-communicative abilities with a university-based ICAP: Aphasia House. Poster presented at The 2016 American Speech-Language-Hearing Association (ASHA) Convention, Philadelphia, PA, USA; 2016.

15. Oelschlaeger ML, Thorne JC. Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. Journal of Speech, Language, and Hearing Research. 1999;42:636-648.

16. Richardson JD, Dalton SG. Main concepts for three different discourse tasks in a large non-clinical sample. Aphasiology. 2016;30 (1):45-73.

17. ASHA: American Speech-Language Hearing Association: Evidence-Based Practice. Available from: http://www.asha.org/Research/EBP/[Accessed 13th March 2016].

18. Kim WS, Kim SH. Research trends of language and communication in the elderly. Journal of Speech-Language & Hearing Disorders. 2017;26(3):35-48.

19. Sung JE, Kim H. Speech and language services in South Korea for individuals with neurogenic communication disorders. Perspectives on Global Issues in Communication Sciences and Related Disorders. 2014;4(1):20-29.

20. Kim H, Na DL. Korean version-Boston Naming Test. Seoul: Hakji-sa; 1997.

21. Kim H, Na DL. Normative data on the Korean version of the Boston Naming Test. Journal of Clinical and Experimental Neuropsychology. 1999;21:127-133.

22. Kim H, Na DL. Normative data on the Korean version of the Western Aphasia Battery. Journal of Clinical and Experimental Neuropsychology. 2004;26(8):1011-1020.

23. Kim H, Na DL. Korean version-the Western Aphasia Battery-Revised. Seoul: Paradise Institute for Children with Disabilities; 2012.

24. Park HS. Korean test for differential diagnosis of aphasia. Seoul: Yonsei University Press; 2006.

25. Kim H, Heo JH, Kim DY, Kim JW. Screening test for aphasia and neurologic-communication disorders (STAND). Seoul: Hakjisa; 2009.

26. Jin C, Choi H, Lee JY. Usefulness of spontaneous speech analysis scales in patients with mild cognitive impairment and dementia of Alzheimer's type. Communication Sciences and Disorders. 2016;21(2):284-294.

27. Kim YN, Sung JE. Story retelling analyses as a function of visual cues using information units for persons with aphasia. Communication Sciences and Disorders. 2017;22(4):756-771.

28. Choi H. Ratio of Correct Information Unit and cognitive functions in healthy elderly adults. Communication Sciences and Disorders. 2015;20(3):435-445.

APPENDIX A

Target main concepts and acceptable alternative lexical items in the English version of MCA Picture Set 1 (Cooking in a kitchen)

The main verb for each main concept is bolded. All the essential information is underlined.

1	The old lady is **cutting** up carrots
2	The old lady **cuts** her finger
3	The old lady's finger is **bleeding**
4	The old lady is **looking** for something in a first-aid box
5	The old lady is **sticking** a Band-Aid

Lexical items that are commonly accepted as alternatives:

Old lady	(older woman, grandma, little old lady, elderly woman, elderly lady, old woman, granny, grandmother, older lady, lady, woman, grandmotherly lady)
Cutting	(slicing, chopping, dicing)
Carrots	(turnips)
Finger	(hand)
Bleeding	(dripping blood)
Looking for	(searching, reaching in)
First-aid box	(medicine cabinet, first-aid kit)
Sticking	(putting, applying, bandaging, afixing, wrapping, placing, using)
Band-Aid	(bandage)

APPENDIX B

Proposed main concepts in Korean and acceptable alternative lexical items for MCA Picture Set 1 (Cooking in a kitchen)

The main verb for each main concept is bolded. All the essential information is underlined.

1	The <u>old lady</u> is **cutting** up <u>carrots</u> <u>할머니가 당근을 **썰고 있어요**</u>.
2	The <u>old lady</u> **cuts** her <u>finger</u> <u>할머니가 손가락을 **베었어요**</u>.
3	The <u>old lady's</u> <u>finger</u> is **bleeding** <u>할머니 손가락에서 **피가 나요**</u>.
4	The <u>old lady</u> is **looking for** something in a <u>first-aid box</u> <u>할머니가 구급상자에서 무엇인가를 **찾고 있어요**</u>.
5	The <u>old lady</u> is **sticking** a <u>Band-Aid</u> <u>할머니가 밴드를 **붙이고 있어요**</u>.

Proposed lexical items as alternatives:

할머니 Old lady	(나이 든 여성, 아주머니, 노인, 노파)
베었어요 Cutting	(자르고 있어요, 다듬고 있어요)
당근 Carrots	(무)
손가락 Finger	(손)
피가 나요 Bleeding	(피가 떨어져요)
찾고 있어요 Looking for	(꺼내고 있어요, 만지고 있어요)
구급상자 First-aid box	(구급함, 구급통, 응급 상자, 약상자)
붙이고 있어요 Sticking	(붙여요, 감싸고 있어요)
밴드 Band-Aid	(반창고, 대일밴드)