Toward true ASL dictionaries: New developments in handshape similarity

Karen Alkoby
DePaul University
School of Computer Science
Chicago, IL
alkobykrn7@rcn.com

Abstract

Despite the oralism movement, ASL was discovered to be a natural language in the 1960s. Previous research has arrived at the consensus that handshape is one of the phonemes of ASL and one of the most recognizable. However, there is little consensus about the number of distinct handshapes in ASL and there is even less consensus about similarity among handshapes. Understanding handshape similarity is important because it is an essential building block toward the development of true ASL dictionaries. True ASL dictionaries would permit people to look up definitions of signs based on ASL phonemes without the use of English. Similarly, using handshapes to look up a sign is essential to the development of true, bilingual ASL/English dictionaries, where it is just as possible to find the English equivalent of a sign as it is to find the sign corresponding to an English word.

Work on handshape similarity initially blossomed in the 1970's, but withered without producing consistent results, due in part to the costly nature of the research. Previous studies involved human subjects performing a large number of comparisons to measure human perception of handshapes. Another impediment was the lack of a realistic 3D representation that respected the natural physiology of the human hand.

In contrast, an analytic model of handshape similarity predicts human perception of handshapes based on the physical positioning of the hand joints in space. Once definitively established, an analytic model has the potential to determine the similarity of additional handshapes without needing to conduct costly perceptual tests.

This paper discusses the creation and initial evaluation of an analytic model of handshape similarity. Initial results are very promising. The model grows more accurate as handshapes are added to it.

1. Handshape

Handshape is the most apparent and complex parameter of a sign and it is the first one that people think of handshape when preparing to make a sign [1, 12]. The handshapes in ASL are composed of letter handshapes, number handshapes, and classifier handshapes [3].

As of today, there is no consensus about the number of distinct handshapes in ASL. Determining handshape similarity is a closely related problem to the determination of the number of handshapes.

Work on handshape similarity initially blossomed in the 1970's [5, 10], but withered without producing consistent results, due in part to the costly nature of the research. Previous studies involved human subjects performing a large number of comparisons to establish human perception of handshapes. Another impediment was the lack of a realistic 3D representation that respected the natural physiology of the human hand.

2. Issues and Benefits

ASL was recognized as a language in 60's and the study of ASL as a language is still a nascent discipline. The current inconsistencies and controversies in handshape usage may have been caused by the oralism movement and the use of various manual codes of English. There are numerous challenges in handshape studies that need to be pushed into next level. These include the problem of expanding a search handshape query when an exact match will not work and determining how to compensate for errors in user's memory recall of a handshape.

A tool that would help in solving these and other questions in handshape studies would be an analytic model of handshape similarity that computes a mapping of the physical positioning of the hand joints in space to human perception of similarity. This paper describes the creation and initial evaluation of an analytic model of handshape similarity. The benefit of such a model is that, once established, it can predict the similarity of additional handshapes without needing to conduct costly perceptual tests. It is an essential building block toward the development of true ASL dictionaries.

3. Plan of Work

This section is an outline of the plan for developing a methodology for computing the similarity of selected ASL handshapes based on the physical (three-dimensional) aspects of handshape production, namely joint rotation. Achieving this will require five steps:

1. Card Sort I, to establish a baseline of human perception of a group of handshapes
2. Model Development, to analyze the card sort data and create candidate models,
3. Model Prediction, where each model computes the similarity of a new, larger set of handshapes,
4. Card Sort II to determine the human perception of the larget set of handshapes and
5. Model Evaluation, which looks at how well each model predicted human perception.

The remainder of the paper discusses each of these steps in detail followed by results and a discussion of future work.

4. Card Sort I

This step establishes a new baseline through user tests and establishes a perception-based similarity among an initial set of handshapes. This step involves card sorting sessions using the same group of handshapes that previous researchers used [5, 10]. Card sorting offers several advantages, including being simple, relatively cheap and quick to execute [7]. It can help to discover users' mental model of an information space. It is an excellent way to gather users' perspective on the handshapes to help classify objects in terms of human perception.

Having 30 partcipants is the minimum for getting a reasonably consistent fit. [11]. Because Stungis had indicated that Deaf and hearing people seem to perceive handshape differently in terms of producing handshapes, this initial card sort interviews 30 Deaf participants and 30 hearing participants.

Thirty Deaf participants completed card sorting sessions at Deaf Expo on October 8, 2005 in Chicago, Illinois. Thirty hearing participants were recruited from students in the ASL/English Interpretation Program at Columbia College Chicago. They completed card sorting sessions during the first week of April 2006. This study passed the DePaul University Institutional Review Board #KA080405CTI-R1.

The session began with obtaining informed consent. Deaf participants viewed the informed consent in ASL on a videotape and hearing participants read the information as written English text. Each participant then completed a very short background questionnaire.

The actual session started with a very short practice session to become familiar with the card-sorting technique. The participants were instructed to sort the cards into piles based on similarity and were free to make as many or as few piles as she or he saw fit. After the practice session, the participant sorted the 20 handshapes. See Appendices A and B for the practice cards and actual cards.

5. Model Development

Model Development occurred in two parts. The first part identifies statistical measure(s) that consistently and accurately summarize the similarities in the card sort data. Consistency measures identify the most appropriate statistics for the card sort data.

The next step is to compare the card sort data to previously recorded three-dimensional joint values to develop a method that assigns similarity measures to the three-dimensional handshape data [2].

5.1 Statistical Analysis Methods

The Pearson's correlation between the card sorting sessions at Deaf Expo and the card sorting sessions at Columbia College is r = .86. One notable difference between two groups is that all participants at Columbia College placed almost every card into one of the piles, but many Deaf Expo participants did not place of all the cards into piles.

Two commonly-used statistics that analyze similarity as clusters are Multi-dimensional Scaling (MDS) and Cluster Analysis; these were used in both of the previous studies [5, 10]. Both methods require the use of a distance (or similarity) measure which is created from the raw card sort data. Figure 1 contains the results of using MDS to visualize the Deaf Expo data.

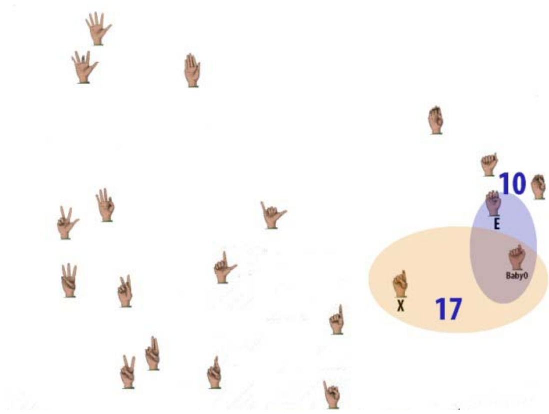


Figure 1: MDS of Deaf Expo

Unfortunately, the MDS of the Deaf Expo data exhibits severe inaccuracies in visualizing the similarities. For example, participants at Deaf Expo grouped "Baby O" and "X" handshapes together into the same cluster 17 times out of 30 times. However, the MDS visualization places the "BabyO" handshape closer to the "E" handshape, which was only paired with "BabyO" 10 times out of 30 times. The stress (measurement of goodness of fit) after several iterations of the MDS procedure was between fair and poor. Therefore, MDS is not an effective statistic for visualizing this data, and it was not used to visualize any additional sets of data.

The second method, Cluster Analysis, displays the data as a tree. See Appendices C and D. In a cluster hierarchy tree, each handshape appears as a leaf node. Handshapes that have a greater similarity are linked at a lower level. Cluster analysis seems more effective for finding handshape similarity in the card sort results, but it is important to determine the goodness of fit. This requires the use of cluster validity measurements. Cluster validity evaluates and assesses the results of a clustering algorithm. There are two types of cluster validity: external and internal criteria. An internal criterion examines how accurately a cluster tree represents the distance in the original data. External criterion examines the amount of agreement between two cluster trees. The cophenetic correlation is as an internal measure [4] and the Rand statistic is an external measure [9].

In a hierarchical cluster tree, any two objects in the original data set are eventually linked together at some level. The level of the connection is the cophenetic distance. Examining the match between the cophenetic distances in a tree and the original card-sort distances is one way to measure how well the computed clustering represents the original distance data.

The Rand statistic is a measure of the goodness of fit between two clustering trees. It computes the probability that two pairs belong to either to a same cluster or to different clusters in both trees [9]. This measurement allows comparison across different levels and number of clusters found within a classification.

The cophenetic correlation for both clusterings were quite high (DeafExpo c = .8924, Columbia c = .9596). So the cluster analysis does accurately summarize the original data. The Rand statistic between the two clusterings was also quite high (R = .84). The high value of the Rand statistic in addition to the high Pearson's correlation would tend to indicate that the Deaf and hearing participants seem to perceive similar patterns in handshape similarity.

5.2 Assigning similarity metrics to joint rotation data

The goal of this second part of model development is to create candidate models for directly computing handshape similarity from physical three-dimensional aspects of the human hand. Twenty joint values are necessary to create a convincing simulation of the poses that the fingers and thumb can assume [8]. Some joint rotates in either abduction-adduction capabilities (spread-close) or flexion-extension motion (bend-open). See Figure 2. The DIP in each finger is highly correlated with the PIP in the same finger. The thumb structure is a little more complex than the fingers but it functions in basicly the same way.

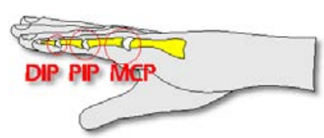


Figure 2: Joints in the fingers

Four candidate models were developed. All four candidate models apply a cluster analysis on either the entire set of handshape data or to a selected subset of it. The entire set of data uses 20 joint values for each handshape. The subset of data also uses joint values, but excludes the DIP of the fingers for a total of 16 joints values. Each handshape pair creates a distance in either 20- or 16-space and these are used in the cluster analysis.

The first two candidate models are Geometric Descriptors which use geometric values. Model 1 uses the entire data set but Model 2 uses the large subset of the values. The last two models are Linguistic Descriptors. These models get their name from the fact that range of geometric data is converted into the range from 0 to 1, analogous to the binary designations used in linguistics. Model 3 uses the entire data set of data but Model 4, the last model derives its values from the selected subset of 16 joints mentioned previously [1, 6].

Tables 1 and 2 summarize the ability of each model to match the card sort data. The cophenetic correlation for each of the candidate models is shown in Table 1. The candidate models that include the DIP joints yield a higher cophenetic correlation.

	Cophenetic Correlation		
Linguistic Descriptors	**With DIP**	**Model 1**	**.7451**
	Without DIP	**Model 2**	**.7211**
Geometric Descriptors	**With DIP**	**Model 3**	**.7383**
	Without DIP	**Model 4**	**.6824**

Table 1. Cophenetic correlations for the candidate models

Table 2 shows the result of computing the Rand Statistic between the clusterings from each of the candidate models and the clusters from the card sorting sessions. Interestingly, omitting the DIP joints yields a slightly better match.

Rand	Geometric Descriptors		Linguistic Descriptors	
	With DIP	**Without DIP**	**With DIP**	**Without DIP**
Combined of Deaf Expo & Columbia	**.7947**	**0.8263**	**.8158**	**.8316**

Table 2. Rand results for the candidate models

6. Model Prediction

In this step, each candidate model will compute (predict) similarity measures from three-dimensional joint data. This step will introduce five additional handshapes to the original 20 handshapes. Each of the candidate methods will compute similarity measures for the 25 handshapes. Figure 2 shows the five new handshapes, which are 'S', 'Vbent', 'Lhook', '8' and '3bent'.

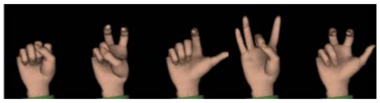


Figure 2. Additional 5 handshapes

7. Card Sort II

The second set of card sorting sessions uses the same procedure as Card Sort I, but includes the additional five handshapes together with initial handshapes. Because the outcome from the first card sessions showed little difference between Deaf and hearing groups, Card Sort II included only Deaf participants. Thirty one interviews with Deaf participants took place during the month of February of 2007 at a Deaf Duppies (Deaf Urban Professionals) event in Chicago and at the Western Suburban Association of the Deaf (WSAD).

8. Model Evaluation

This last stage assesses the predictive validity of the candidate models by comparing the similarity measures computed by the model to the empirical results of Card Sort II. To satisfy the goal of predictive ability, it was necessary obtain improvement in both internal and external criteria when additional handshapes were added. The results from each of the candidate models were compared to the results from Card Sort II using both cophenetic correlation and the Rand statistic. As a final check, the five new handshapes were deleted from the Card Sort II data, reclustered, and compared with the initial card sort sessions.

9. Assessment of the candidate models

For the linguistic descriptors model without DIP joints, the cophenetic correlations improve. However, for the models that use Geometric Descriptors to represent handshapes, the cophenetic correlations indicate that it is better to include all 20 joint angles. See Table 3.

Cophenetic Correlation				
	Geometric Descriptors		Linguistic Descriptors	
	With DIP	Without DIP	With DIP	Without DIP
	Model 1	Model 2	Model 3	Model 4
Card Sort I	.7383	.6824	.7451	.7211
Card Sort II	.7510	.6997	.7218	.7343

Table 3. Cophentic Correlation

The Linguistic Descriptor approach is particularly promising because it shows an efficiency that the Geometric Descriptor approach did not. When the DIP angles were omitted from the Geometric Descriptor, it performed more poorly. But when the DIPs were omitted from the Linguistic Descriptor model, the cophenetic correlation improved, and its Rand statistic (agreement with user perception) was the best of all the models. See Table 4.

Rand Statistic				
	Geometric Descriptors		Linguistic Descriptors	
	With DIP	Without DIP	With DIP	Without DIP
	Model 1	Model 2	Model 3	Model 4
Card Sort I	.7947	.8263	.8158	.8316
Card Sort II	.8633	.76	.8667	.8867

Table 4. Rand Statistic

In Card Sort II, the cophenetic correlation and the Rand statistic improved for the Linguistic Descriptor without DIP joints. Therefore, the Model 4 is the best candidate for predictive ability. This model is particularly promising because it appears that it produces better results when incorporating additional handshapes in the model. However, all of 4 candidate models should still be evaluated and validated with a still larger number of handshapes.

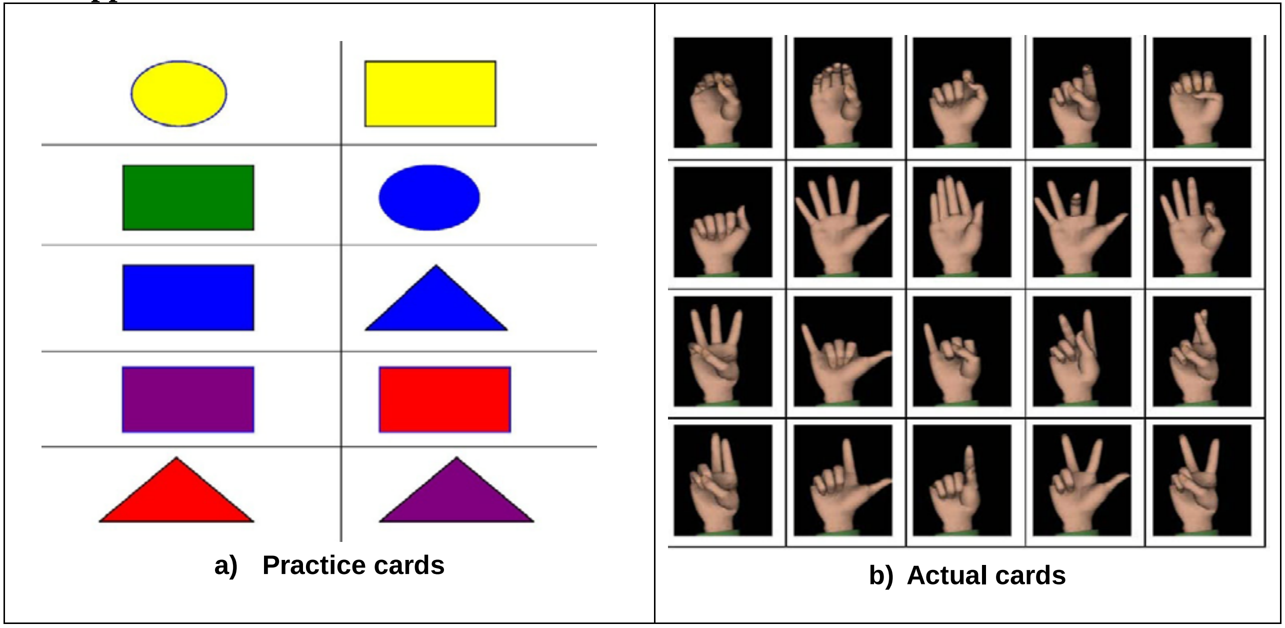
As a check, a final cluster analysis was performed on the data gathered from Card Sort II after the five new handshapes were removed. The cophenetic correlation of the result is extremely high (c = .9202). The Rand statistic between the Card Sort II subset and the combined set of Deaf Expo and Columbia from Card Sort I is .87. This indicates that the addition of new handshapes did not affect the overall perception of handshape similarity.

10. Future Work

Using a cluster analysis with the properly chosen representation appears to produce better results as the number of handshapes increases. A next step would be to introduce additional handshapes to create a large set of approximately 40 handshapes to further test the all of the candidate models.

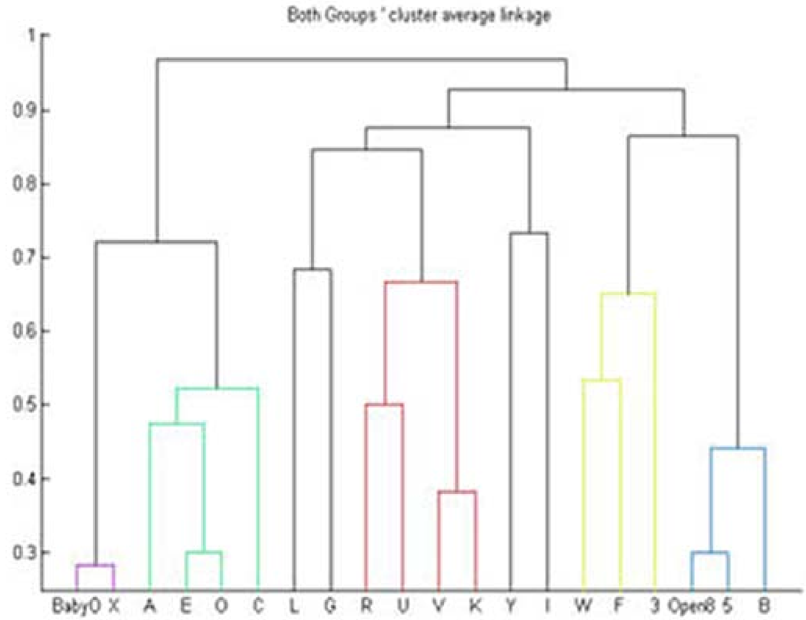
The handshape similarity concept has the potential creating effective computerized ASL/English dictionary searching. Finding a next handshape similarity can be addressed either due to misidentifying or memory error recall which can be rectified in "fuzzy" searches for ASL signs. For example, one may misidentify the handshape for the sign "DOUBT" by selecting "3-bent" handshape, because it is very similar to V-bent. If a search on a "3-bent" handshape yielded no desired results, the user could elect to expand the search by considering similar handshapes. This fuzzy search would look for handshapes that were closest in appearance to the "3-bent" handshape.

11. Appendices

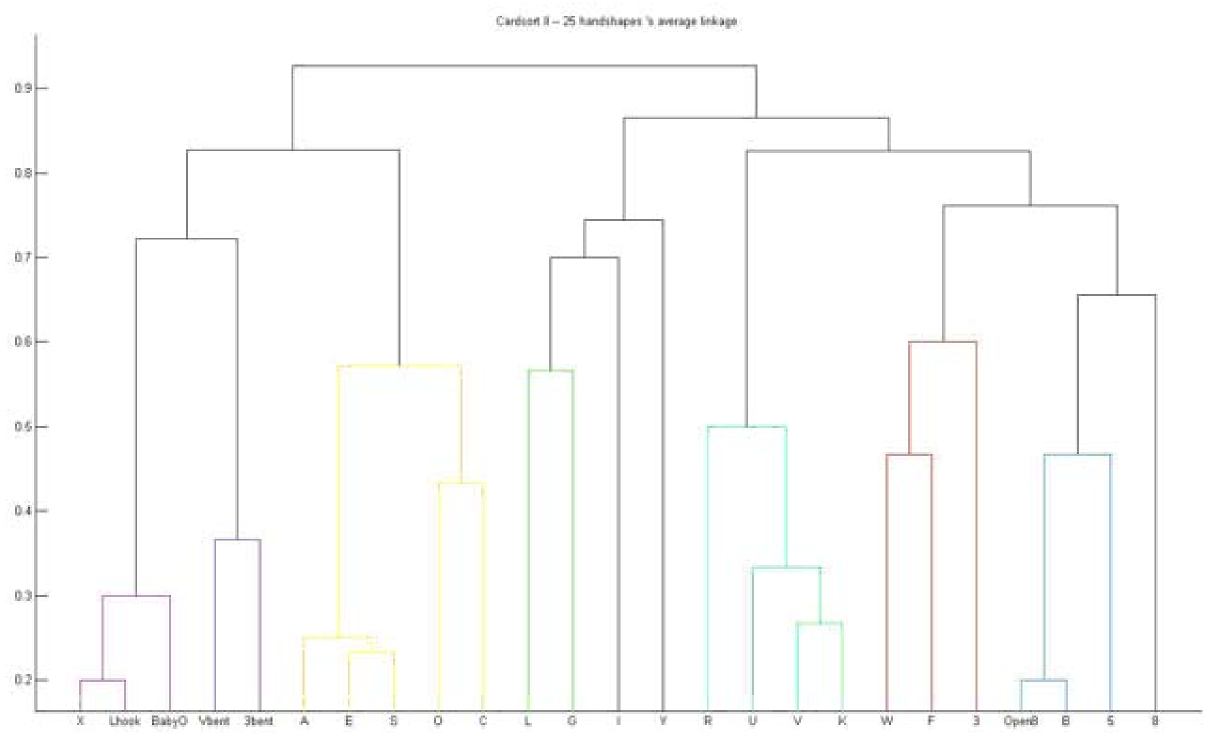


a) Practice cards

b) Actual cards



c) Card Sort I 's Cluster Tree



d) Card Sort II's Cluster Tree

References

[1] *Brentari, D. (1998). A Prosadic Model of Sign Language Phonology. Cambridge, MA: MIT Press.*

[2] *Davidson, M.J., Alkoby, K., Sedgwick, E., Berthiaume, A., Roymeico, C., Christopher, J., Craft, B., Furst, J., Hinkle, D., Konie, B., Lancaster, G., Luecking, S., Morris, A., McDonald, J., Tomuro, N. , Toro, J., Wolfe, R. (2000, November 4, 2000). Usability Testing of Computer Animation of Fingerspelling for American Sign Language, Presented at the 2000 DePaul CTI Research Conference, Chicago, IL.*

[3] *Grieve-Smith, A. B. (1998). Sign Synthesis and Sign Phonology. Paper presented at the Proceedings of the First Annual High Desert Linguistics Society Conference, Albuquerque:HDLS.*

[4] *Halkidi, M., Batistakis Y., and Vazirglannis, M. (2002, June 2002) Cluster Validity Methods: Part I. SIGMOD Record Vol. 31, No. 2, 40-44*

[5] *Lane, H., Boyes-Braem, P. , & Bellugi, U. (1979). A feature Analysis of handshape. In e. Klima & Bellugi (Ed.), The signs of Language (pp. 164-180). Cambridge, MA: Harvard University Press.*

[6] *Liddell, S., & Johnson, R.E. (1989). American sign language: The phonological base. Sign Language Studies, 63, 195-277.*

[7] *Maurer, D., and Warfel, Todd. (Retrieved 04/07/2004). Card sorting: a definitive guide, from http://www.boxesandarrows.com/archives/card_sorting_a_definitive_guide.php*

[8] *McDonald, J., Toro, J., Alkoby, K., Berthiaume, A., Chomwong, P., Davidson, M., Konie, B., Lancaster, G., Lytinen, S., Roychouhuri, L., Furst, J., Sedgwick, E., Tomuro, N. & Wolfe, R. (2000). An Improved Articulation Model of the Human Hand. Proceedings of the ISCA 15th International Conference on Computers and Their Applications (CATA-2000) 427-430.*

[9] *Rand, William. (1971) Objective Criteria for the Evaluation of Cluster Methods, Journal of American Statistical Association. Vol. 66, Number 336 in Theory and Methods Section*

[10] *Stungis, J. (1981). Identification and discrimination of handshape in American Sign Language. Perception and Psychophysics, 29, 261-276.*

[11] *Tullis, & Wood (2004, June 7-11, 2004). How Many Users Are Enough for a Card-Sorting Study? Paper presented at the Proceedings UPA 2004', Minneapolis, MN.*

[12] *Valli, C., & Lucas, C. (1995). Linguistics of American sign language : An introduction (2nd ed.). Washington, D.C.: Gallaudet University Press.*