Change in Word Fluency over Adulthood: A Longitudinal Linguistic Cluster Analysis

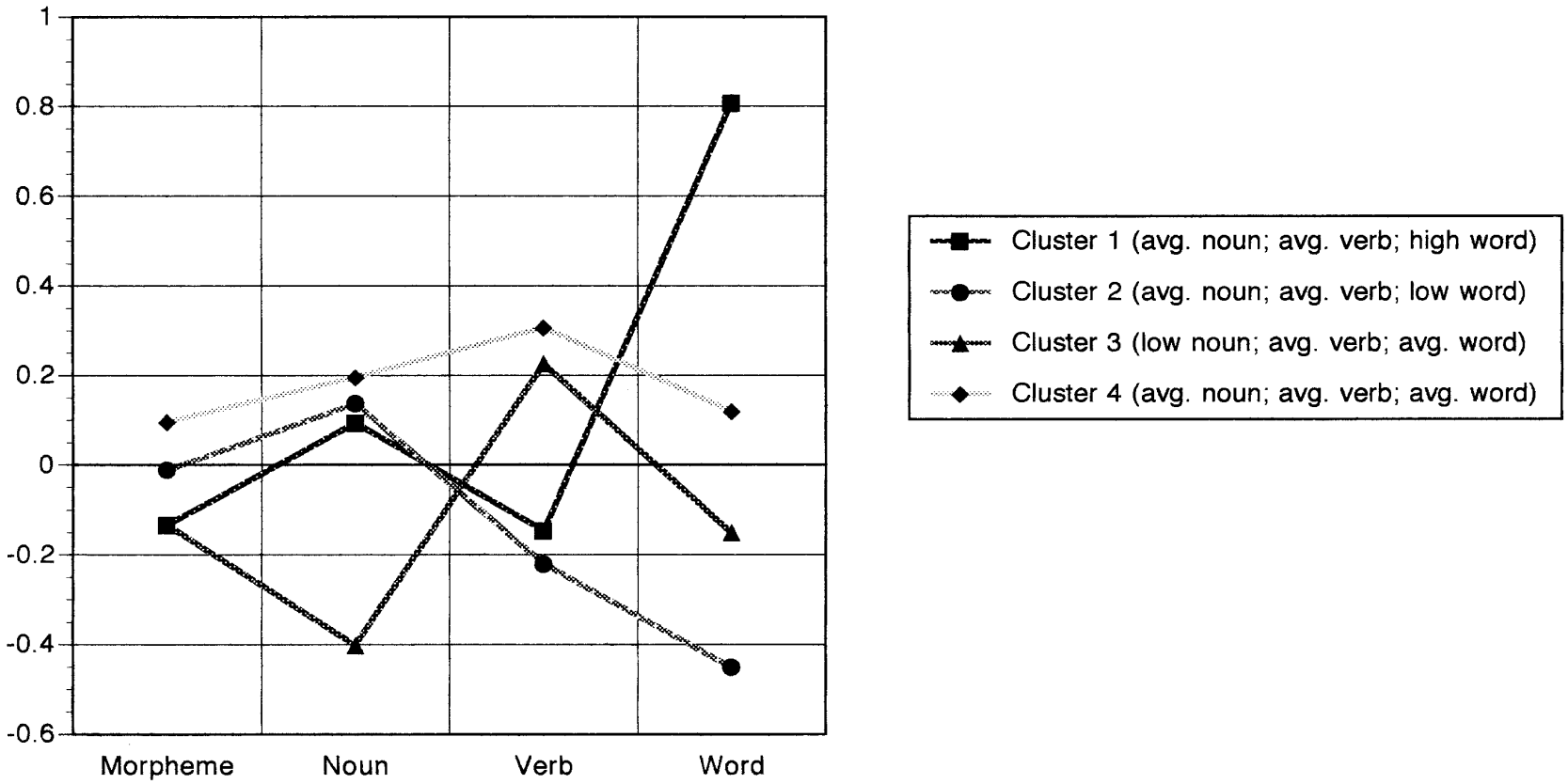
Megan M. Tesluk, K. Warner Schaie, and Sherry L. Willis

The Pennsylvania State University
Human Develoment and Family Studies
115 South Henderson Building
University Park, PA 16802

Running Head: WORD FLUENCY

This research is supported by Grant R37 AG08055 from the National Instituite on Aging to the second author. The authors greatly acknowledge the cooperation of the Group health Cooperative of Puget Sound.

Figure 6: 1991 Sample 2 Data



Abstract

Word Fluency serves as an assessment tool to examine many topics in the study of cognition, such as mental functioning, individual differences, and dementia. Prior research on word fluency has been conducted using the total number of responses provided within a fixed time period. Most word fluency data are cross-sectional. This study contributes to the literature by examining word fluency from a linguistic perspective involving grammatical and morphological analyses with a hierachical cluster analysis. The study examined longitudinal change and qualitative differences in word production. Four distinct patterns of preformance (clusters) emerged at both time periods. Replication was high within time periods, but was not as high across time. Finally, cluster membership was not found to be stable or consistent across time.

Many researchers have used the Word Fluency Test (Thurstone & Thurstone, 1948) as an assessment tool for examining areas of cognition such as mental functioning, individual differences, and dementia in adults. Traditionally, word fluency has been evaluated cross-sectionally by analyzing the number of words produced quantitatively. It has been reported that there are gender differences favoring women in word frequency tasks (Alvis, Ward, & Dodson 1989; Buckelew & Hannay 1986). Additional research demonstrates that word fluency scores decline with age (Alexander, Langer, Newman, Chandler, & Davies 1989; McCrae, Arenberg, & Costa 1987; Obler & Albert, 1985).

An alternative approach to the quantitative analysis of word fluency involves a qualitative linguistic perspective. Linguistic analysis can be approached grammatically or morphologically. Grammatical analysis involves categorizing words into their parts of speech (e.g., nouns, and verbs), while a morphological analysis characterizes words into the appropriate number of morphological parts. These methods examine not only the number of responses produced but also the types or categories of responses that subjects produce. Qualitative changes in word production can be measured using several language characteristics as opposed to one general score.

A morphological analysis of adult speech is a way to examine the complexity of one's word responses. The unit of analysis, a morpheme, is the smallest meaningful unit in any language. Its form and meaning are consistent throughout a language. Words can consist of either one or several morphemes. Words like *tree*, *cat*, *run*, *sad*, *nice*, and *think* all consist of one morpheme, and can not be broken down into smaller lexically meaningful segments. These words are called free morphemes because they can stand alone and express their meaning. The other major type of morpheme is bound and can not express its lexical meaning without being attached to another word or root. Suffixes, prefixes, and roots (i.e., *-ed*, *-ment*, *-ing*, *-ly*, *sub-*, *per-*, *-mit*) can slightly change the meaning of the original word when they are attached to other forms. Newly formed words like

agree+ment, *sad+ly*, and *sub+marine* have two morphemes. Words can contain several morphemes, like *un+happi+ness*, *un+lik+ly+ness*, and *non+pre+condition+al+ly*.

Morphemes are considered as a measure of complexity in child language acquisition studies (Brown, 1973). Morphological analysis is used as a developmental marker to determine if a child has acquired grammatical rules of their language (Berko, 1958). Basically, the more morphemes a word has the more different lexically meaningful parts it contains, and thus the more complex it becomes. The same type of measure can be applied to adult language. The present investigation used the average number of morphemes per word in the analysis of the data.

An earlier investigation (Tesluk, McGuire, Schaie, & Willis; 1994) examined how adults' word frequency responses change over time. The study examined 360 participants from the Seattle Longitudinal Study (SLS). Data were collected at two time periods (1984 and 1991) on all 360 subjects. The examination used variables that had not previously been used with the word fluency task. The study hypothesized that with more detailed measures of performance, age and gender differences would be minimized. However, the results suggested that gender differences were not reduced by showing that women performed better than men at both time points. Age differences were only effected by the morphological analysis. The results suggested that although the number of words produced was assumed to decrease with age; the level of complexity of the words produced increased. Other trends in the data were not as conclusive as the results for the complexity measure. The number of nouns remained essentially the same across time, but verbs showed a significant decrease between time 1 and time 2. No other parts of speech demonstrated significant results.

The present study is the first to examine word frequency responses longitudinally using grammatical and morphological variables with cluster analysis. Cluster analysis is a heuristic tool useful for generating hypotheses. It is also an exploratory methodology useful for examining how people or entities can be grouped together on a specific set of

variables. Previous studies and the experiment cited above have shown that performance (number of responses) decreases with age while the complexity of responses increases. However, no information has been provided about the characteristics or patterns of people's expressions of their lexicon. Cluster analysis detects if common patterns used by participants in word fluency exist, and whether these patterns change over time. This examination of word fluency will help to generate hypotheses concerning different types of lexical expression.

Method

Participants

Participants were subjects from the Seattle Longitudinal Study (SLS). This large scale longitudinal-sequential study of adult cognitive development begun in 1956, has involved more than 5,000 participants (Schaie, 1983; 1993). All participants in the present investigation took part in the 1984 and the 1991 assessment sessions. Participants (N = 360; Males = 154, Females = 206) ranged in age from 42 to 92 years old (age at time 2). The mean educational level was 14.61 years (SD = 3.01, range = 1-20), and the mean income was $ 27,700 (SD = $8,160, range = $ 2,000 - $ 34,00). A description of the sample found in Table 1 further divides participants by gender.

Insert Table 1 about here

Materials and Procedure

Participants were administered the Word Fluency Test from the Primary Mental Abilities battery (Thurstone & Thurstone, 1948) in 1984 and 1991. Word fluency is a task concerned with speeded verbal production. Participants were given five minutes to write down as many words as they could beginning with the letter 's'. Participants were

instructed not to list proper nouns or the conjugation of verbs (i.e., not to include the present and past tense of the same verb) in their responses.

Each word was categorized by parts of speech and by number of morphemes. The parts of speech were determined by The English Oxford Dictionary (1973). Two independent raters trained in Linguistics divided the words into morphemes. When the raters did not agree on the morphological division, the etymology of the word was consulted to determine which morpheme value should be used. Raters disagreed on only 5 % of the corpus of responses.

Results

Tables 2 and 3 present by age and gender the number of words produced, the average number of morphemes per word, and the percentage of nouns, verbs, adjectives, and adverbs. The number of adjectives, adverbs and prepositions consisted of such a small portion of subjects' responses that they were omitted from the cluster analysis.

Insert Tables 2 and 3 about here

Cluster Analysis

The data were analyzed using the average linkage method, with a cosine measurement. Although several different methods are available for cluster analysis, the average linkage cluster analysis method was selected because it is considered to be a compromise between the problems with single linkage and complete linkage method (Blashfield, 1976; Sokal and Sneath, 1963). Average linkage method is defined as "a group of entities in which each member has a greater mean similarity with all members of the same cluster than it does with all members of any other cluster" (Blashfield, 1976, p. 379). The cosine measure was selected because it includes the shape, but also the level and scatter of the cluster in its measure.

Subjects were clustered by the following variables: number of words, portion of words that are nouns, portion of words that are verbs, and average number of morphemes per word. Nonwords and errors were not eliminated from the data. Therefore it is possible for a cluster profile to have a high portion of verbs and nouns, while only having an average number of words. The profile might also contain a low number of other parts of speech (adjectives and adverbs) or errors.

The sample of 360 subjects was randomly split into two subsamples (N=190 and N=170) for replication purposes. After the clusters were created, one-way ANOVAs were run to determine whether there were significant differences between the clusters' mean scores on each of the four variables. Tukey's honestly significant difference test (HSD) were used to compute the differences between clusters on a given variable. This test was chosen for its way of dealing with unequal cell (cluster) sizes. All scores reported on the word fluency task were standardized (z scores). This was necessary in order to demonstrate meaningful differences between clusters and for cluster assignment.

Presence of clusters in the data

The cluster analysis results revealed that the data have a five cluster solution for both time periods. Four distinct patterns are present for time 1 (1984) and time 2 (1991), and a small cluster of outliers was present in each sample (See figure 1). The same clusters were found when the sample was randomly split and replicated for both time periods. The stopping rule involved with the clustering method specified to stop clustering when the newly formed clusters were either less than ten percentage of the total population, or the new cluster did not add any new information to the solution. The replication samples for both time periods provided the same number of clusters (4), with a group of outliers (see figure 2).

Insert Figures 1 and 2 about here

Composition of the clusters

Each cluster had a distinct pattern. The four clusters from the 1984 data are described in Figure 3. Replication of the sample is presented in figure 4. One cluster had a profile depicted by an average number of words, containing a low portion of nouns and an average portion of verbs. A second cluster had a profile of an average number of words containing a higher percentage of nouns and verbs. A third cluster profile was characterized by a high number of words, with a high percentage of nouns and low percentage of verbs. The final cluster from the 1984 data was characterized as having low number of words, with a high number of nouns and a low percentage of verbs. One-way Anovas performed on the clusters showed that there were significant differences across clusters on verbs, nouns, and number of words. The morpheme score showed no variation between clusters. The same pattern of significant differences was found in the replication. Tukey tests revealed that cluster one had significantly more nouns than the other clusters. Only clusters 3 and 4 did not differ significantly on verbs. Cluster one was significantly different from the other in the total numbers of words produced.

Insert Figures 3 and 4 about here

In the 1991 data, four distinct clusters emerged again (see figure 5). The replication of this sample is presented in figure 6. Two clusters had an average number of words and an average-high percentage of verbs, but each differed in the percentage of nouns. One cluster had an average percentage of nouns and the other had an average-high percentage of nouns. A third cluster had a high number of words which contained an average percentage of both nouns and verbs. The fourth cluster contained an average-low number of words, with an average-low percentage of both nouns and verbs. The ANOVA results for these clusters showed that for nouns and the number of words there were significant differences between clusters. In the replication sample, nouns, number of words, and verbs were

significantly different. Morphemes had variation, and differed significantly. Tukey tests showed a significant difference between clusters 3, 4 and 1,2 on the total number of words produced. Clusters 1 and 3 differed significantly from 2 and 4 on the number of morphemes. All clusters were significantly different from each other on nouns. The same was true for verbs except for clusters 1 and 3. The same general pattern pertaining to cluster names applied to the 1991 data as well.

Insert Figures 5 and 6 about here

Similarity between replications and over time

Q-correlation tests were used to asses the similarity in cluster shape between the study and replication samples. Comparing the 1984 samples the Q-correlations ranged from .999 to .924. The Q-correlation for the 1991 samples ranged from .994 to .940 for three clusters. The fourth cluster had a lower Q-correlation of .701. The 1984 data appears to have selected clusters with the same shape. The 1991 data selected at least three clusters with the same shape. The fourth cluster matched moderately well.

Q-correlations were also used to test the similarity in shape over time. The 1984 sample 1 clusters were correlated with those from the 1991 sample 1. Two clusters matched their shape well, with correlations of .872 and .825. One cluster reported a modest similarity with a correlation of .684. The final cluster had no relation in shape between the 1984 and 1991 samples with a correlation of -.446. However, the replication samples showed much more similarity in the shape of its clusters. The Q-correlations ranged from .990 to. 802 for all the clusters.

Membership within clusters across time periods

A reasonable next question to ask is whether the membership of these four clusters changed over time; and if so, were there distinguishable patterns. After a close examination, it was determined that the people in each cluster changed from time1 to time2. Pairs of membership were listed by subject. For any one cluster in 1984, some of its members moved to each cluster in 1991. There were some (mildly strong) patterns of change. The most common patterns were for 'average Joes' in 1984 to become 'quick thinkers'. Many of the 'direction followers' of 1984 became 'average to low Joes' in 1991. Many 'slower thinkers' in 1984 became 'average Joes' in 1991. A majority of the 'quick thinkers' in 1984 became members of two different clusters at time 2. They either became 'average Joes' or 'direction followers'.

Assignment rules were not possible for patterns over time. This was largely due to the developmental nature of the data, and the failure to replicate similarity patterns with the

Q-correlation. It was expected that subjects would change in performance. Assignment rules for the 1984 data had a coverage rate of about 76% for both samples . Each rule had to be general enough to capture the strongest characteristic of each cluster. The rules were written with standardized scores so that they would better generalize to other samples.

Insert Table 4 about here

Assignment rules for the 1991 data were possible to formulate. Patterns of significant differences between variables and groups were not large enough or stable enough (over replication) to allow for assignment rules even with the most modest amount of coverage. A major reason for this inability could be due to the reduced variation of the cluster means on the variables.

External correlates

Because age and gender were important variables in previous studies of word fluency, it was hypothesized that they would be meaningful external correlates to this analysis. Cross tabulations were performed with age and gender on the 1984 and 1991 samples. No meaningful patterns emerged. No cell had an unusually high or low number of members, which would have provided a differentiating characteristic of that cluster.

Discussion

The previous research on the Word Fluency task has provided quantitative information including that the number of responses decreases with age and that women perform better than men on the task. The previous investigation (Tesluk, et al., 1994) used a qualitative approach, examining specifically changes in grammar and word complexity that occur as participants age. The data revealed that with age the number of verbs decreases. The decrease in the number of verb responses could be due to the lexical rule

placed on verbs (no conjugations). A lexical rule such as no proper nouns effects a smaller portion of the noun category than the portion effected by the lexical restriction on verbs. This could make the category of verbs more limited or difficult to produce. Additional research needs to be conducted to examine whether this difficulty increases with advancing age. The previous examination also found that the complexity of words generated (i.e., number of morphemes per word) increases with age.

The significance of the current investigation pertained to patterns of linguistic change on a cognitive task. The use of cluster analysis lent support the idea that people's lexical expressions differ based on the number of words and categories they can generate under a certain time period. The results showed that there were four types of patterns for performance. The same result was found for the same group of subjects seven years later. However, there was no clear pattern of change or stability across time for these subjects.

The establishment of clusters at both time periods suggests that people do have either lexicons that are arranged differently or different word retrieval patterns. Possibly some subjects' lexicons include more listings of parts of words and nonwords. This could help to explain the high number of words recalled, but allowing for an average or low numbers of nouns or verbs in some subjects' performances. Some subjects might recall verbs more easily than nouns, explaining the high percentage of verbs in their answer.

The use of cluster analysis generated three possible hypotheses about changes in performance over time. The results suggest several explanations for no clear pattern of change. Possibly people change in random patterns. Or more likely there are patterns, but the sample of 360 was too homogeneous to detect the differences. The final possible explanation would concern fewer number of clusters. This data set might have picked four patterns for each time, but possibly there are only two general patterns of change.

It was surprising that the morpheme variable showed no significant differences between cluster groups. It was a significant factor in the previous experiment. A possible explanation was that the frequency was highly positively skewed. A suggestion for future

analysis would be to rerun these analyses after performing transformations to improve the quality of the morpheme distribution.

Several other results that emerged suggested that a larger data set might be helpful for determining patterns over time. The addition of one or two other verbal recall tests that contain the same variables (nouns, verbs, morphemes, and total number of words) maybe instrumental for distinguishing patterns of performance over time. The same variables used with different tasks may tap into people's differentiating abilities and more importantly how those variables change over time.

References

Alexander, C., Langer, E., Newman, R., Chandler, H., & Davies, J. (1989). Transcendental meditation, mindfulness, and longevity: An experimental study with the elderly. *Journal of Personality and Social Psychology*, *57*, 950-964.

Alvis, G., Ward, J., & Dodson, D. (1989). Equivalence of male and female performance on a tactuospatial maze. *Bulletin of Psychonomic Society*, *27*, 29-30.

Berko, J. (1958). The child's learning of English morphology. *Word ,14*, 150-177.

Blashfield, R. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychology Bulletin*, *83*, 377-388.

Brown, R. (1973). *A first language: The early stages*. Cambridge, Mass. Harvard University Press.

Buckelew, S. & Hannay, J. (1986). Relationships among anxiety defensiveness, sex, task difficulty, and performance on various neuropsychological tasks. *Perceptual and Motor Skills*, *63*, 711-718.

McCrae, R., Arenberg, D., & Costa, P. (1987). Declines in divergent thinking with age: Cross-sectional, longitudinal, and cross-sequential analyses. *Psychology and Aging*, *2*, 130-137.

Obler, L. K. & Albert M. L. (1985). Language skills across adulthood. In J. E. Birren & K. W. Schaie (Eds.) *Handbook of the Psychology of Aging* (Second Edition) (pp. 463-473) New York. Van Nostrand Reinhold.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64-135) New York: Guilford.

Schaie, K. W. (1993). The course of adult intellectual development. *American Psychologist, 49.* pp.1-16.

Sokal, R. & Sneath, P. (1963). *Principles of numerical taxonomy*. San Francisco: W. H. Freeman.

Tesluk, M., McGuire, C., Schaie, K.W. & Willis, S. (1994). Change in word fluency over the adult lifespan: A longitudinal linguistic analysis. Poster presented at Cognitive Aging Conference, Atlanta, Ga. April.

Thurstone, L. L. & Thurstone, T. G. (1949). *Examiner manual for the SRA Mental Abilities Test* (Form 10-14). Chicago: Science Research Associates.

Table 1.

Sample by gender and age cohort.

	Middle-Aged			Young-Old			Old-Old		
	M	F	T	F	M	T	M	F	T
N	39	49	88	91	98	189	27	56	83
Age	56.46	55.87	56.08	70.30	70.70	70.52	81.18	82.02	81.66
SD	5.90	5.77	5.80	4.01	3.75	3.86	3.70	3.41	3.53
Range	43-63	43-63	43-63	64-77	64-77	64-77	78-92	78-92	78-92
Education	14.58	14.35	14.43	15.31	13.85	14.53	14.79	13.51	14.12
SD	3.35	2.35	2.77	2.64	2.36	2.59	4.05	2.81	3.43
Range	4-20	11-20	4-20	10-20	8-20	8-20	1-20	8-18	1-20

Table 2.

Language Variables (Frequency): Age Cohort, Gender, Occasion

	Middle-Aged				Young-Old				Old-Old			
	1984		1991		1984		1991		1984		1991	
	M	F	M	F	M	F	M	F	M	F	M	F
X # Words	43.84	45.46	43.34	47.39	42.24	39.84	40.44	40.02	34.00	39.36	31.13	38.73
SD	12.28	12.70	11.85	12.28	11.59	12.24	12.86	12.69	11.96	12.15	13.28	12.87
Range	23-77	14-80	22-67	20-80	13-71	4-75	13-79	8-72	14-70	16-62	9-67	10-67
X Morphemes /per word	1.12	1.15	1.17	1.18	1.16	1.18	1.16	1.18	1.30	1.14	1.70	1.16
SD	.08	.09	.11	.10	.10	.10	.10	.11	.09	.10	.11	.11
Range	1-1.31	.98-1.40	1.03-1.58	1-1.53	.93-1.50	1-1.62	.95-1.50	1-1.75	.88-1.29	1-1.38	1.03-1.50	1-1.39

Note. M = male; F = female.

Table 3.
Language Variables (Percent); Age Cohort, Gender, and Occasion.

	Middle-Aged				Young-Old				Old-Old			
	1984		1991		1984		1991		1984		1991	
	M	F	M	F	M	F	M	F	M	F	M	F
% Nouns	.56	.55	.55	.56	.56	.55	.56	.55	.54	.55	.56	.56
SD	.09	.10	.09	.09	.10	.11	.11	.11	.09	.10	.14	.12
Range	.41-.81	.14-.79	.37-.78	.35-.80	.36-.90	.32-.81	.16-.93	.34-.88	.33-.72	.31-.76	.22-.74	.30-.80
% Verbs	.44	.43	.45	.44	.42	.42	.43	.42	.36	.43	.42	.42
SD	.10	.10	.10	.10	.11	.12	.10	.12	.10	.12	.13	.12
Range	.19-.70	.20-.68	.21-.72	.21-.68	.14-.63	.16-.62	.07-.65	.13-.82	.08-.50	.10-.61	.08-.62	.13-.67
% Adjectives	.17	.17	.18	.18	.18	.18	.18	.18	.19	.19	.22	.18
SD	.07	.08	.08	.06	.08	.08	.09	.10	.08	.09	.11	.09
Range	.03-.33	.03-.44	.04-.42	.05-.33	0-.40	0-.40	0-.50	0-.50	0-.07	.03-.51	0-.10	.04-.46
% Adverbs	.01	.01	.01	.02	.01	.02	.02	.02	.02	.02	.02	.01
SD	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.03	.02
Range	0-.06	0-.06	0-.09	0-.08	0-.08	0-.10	0-.11	0-.08	0-.07	0-.06	0-.10	0-.09

Note. M = male; F = female.

Table 4.
Assignment Rules

If the noun score is low (below 0),
verb score is close to 0,
the word score is close to 0..... .....Assign to Cluster 1

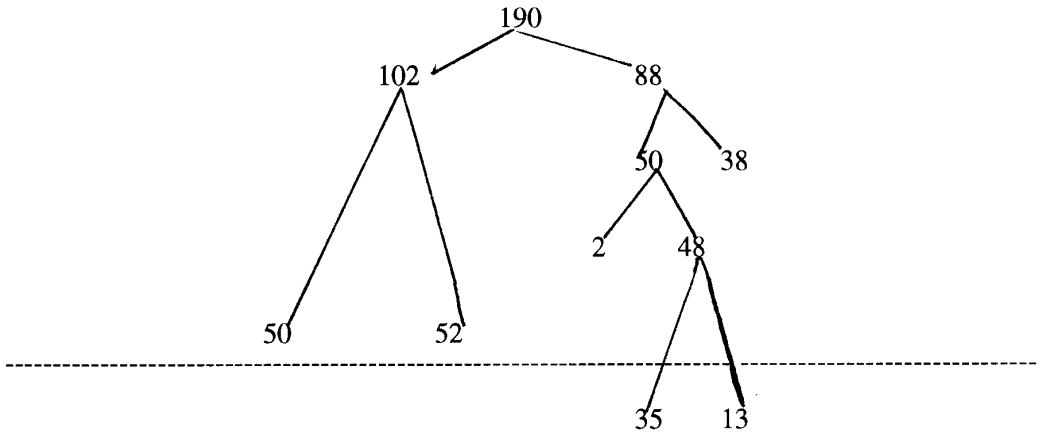
If the noun score is high (above 0),
verb score is hogh (above 0),
the word score is close to 0..... .....Assign to Cluster 2

If the noun score is close to 0,
verb score is close to 0,
the word score is high (above 0)..... ....Assign to Cluster 3

If the noun score is close to 0,
verb score is close to 0,
the word score is low (below 0).....Assign to Cluster 4

Figure 1.
Tree diagram of clustering.

Sample 1 1984 data



Sample 1 1984 data

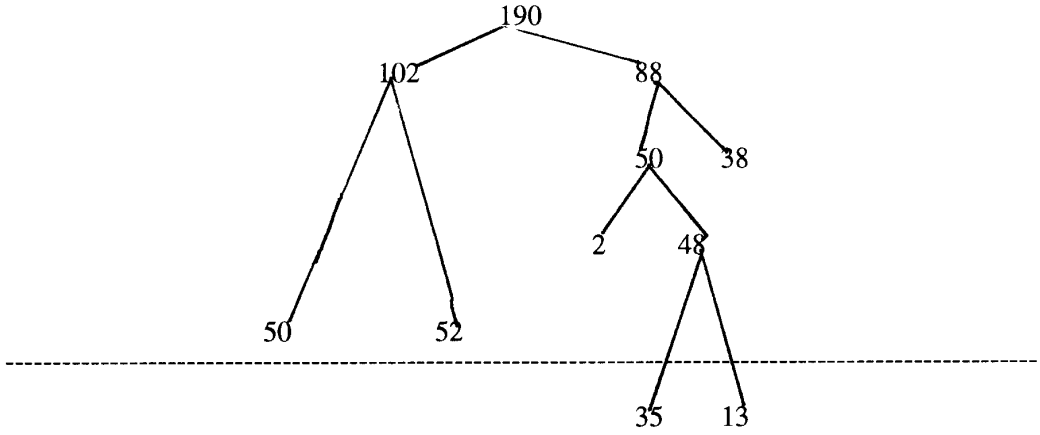
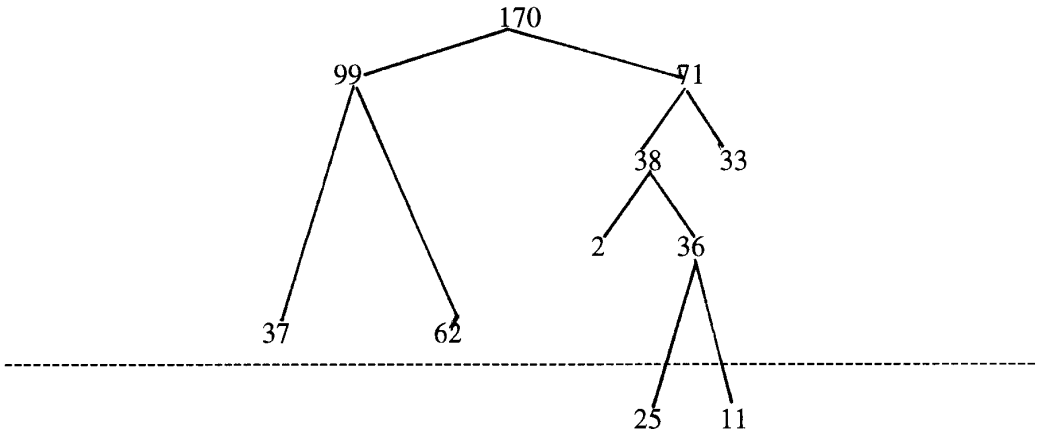


Figure 2.
Tree diagram of clustering.

Sample 1 1984 data



Sample 1 1991 data

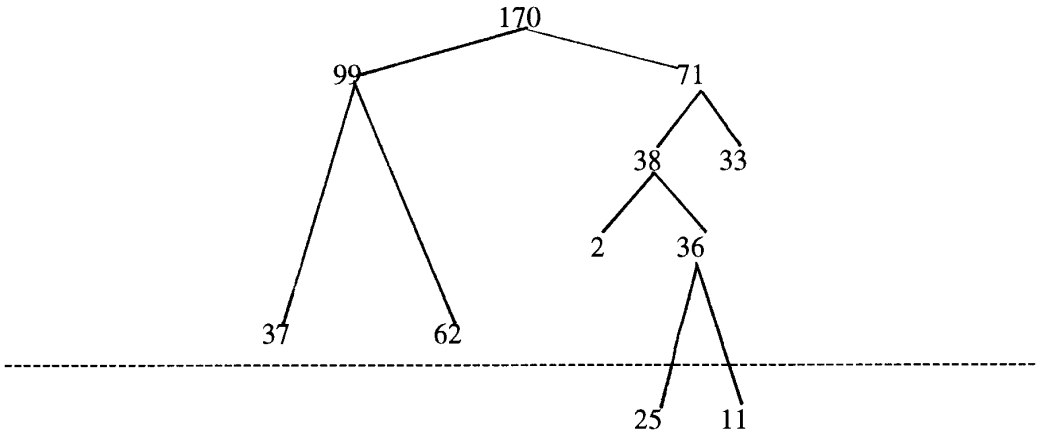


Figure 3: 1984 Sample 1 Data

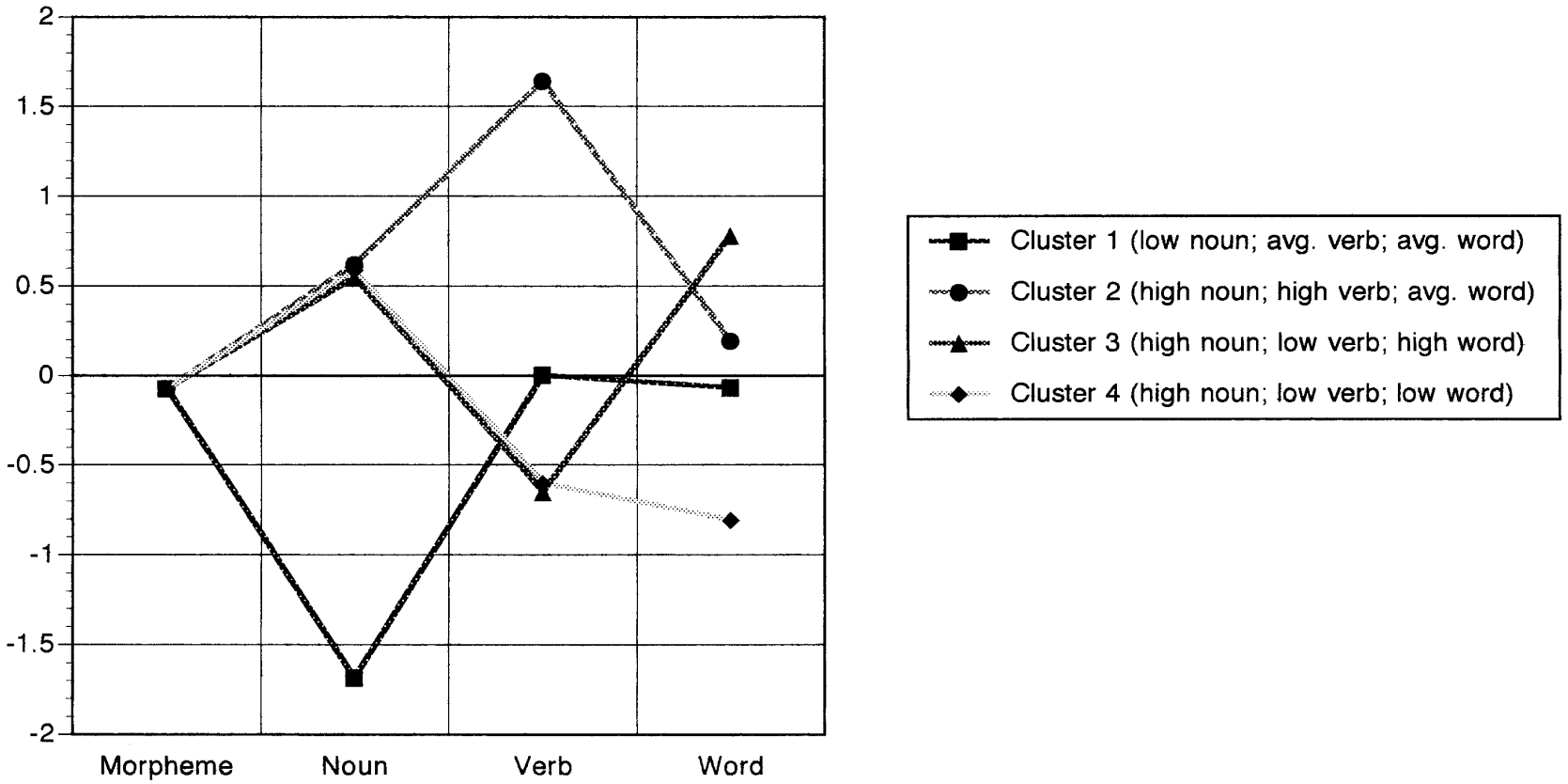


Figure 4: 1984 Sample 2 Data

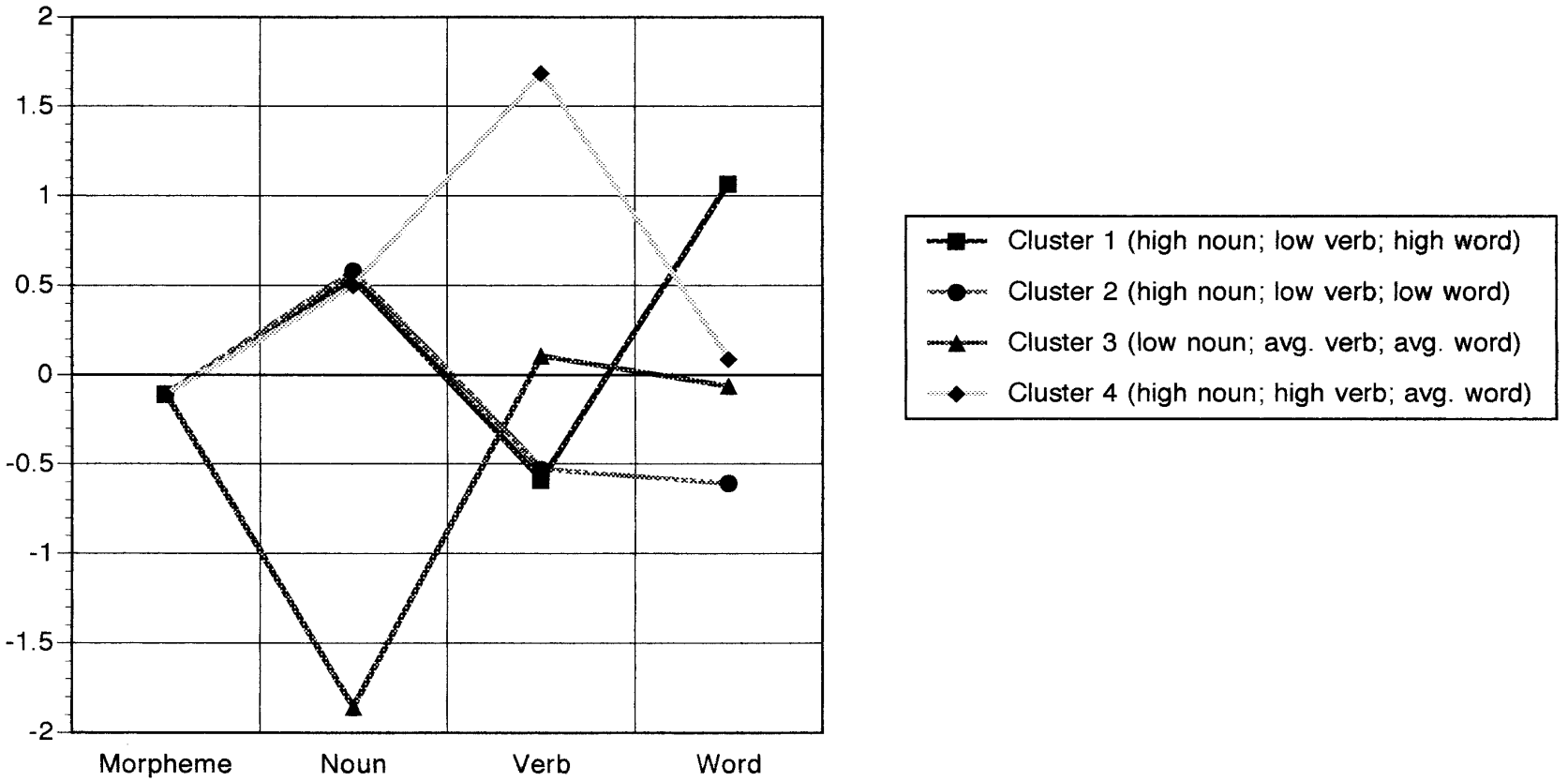


Figure 5: 1991 Sample 1 Data