

# CONSIDERATIONS WHEN CONSTRUCTING A SEMANTIC DIFFERENTIAL SCALE

Jayne Al-Hindawe

## Abstract

*The vast majority of language attitude studies have employed the use of a semantic differential scale to elicit the listener's perceptions of various language varieties and the speakers of those varieties. However, after administering such a scale in a small pilot study it became apparent that there were numerous unforeseen decisions which had to be made in designing the scale. The process of designing a semantic differential scale has been rarely discussed in the literature. This process can essentially be broken down into two areas; the selection of adjectives and the layout of the scale itself. Decisions made regarding these two factors require serious thought if results are to be meaningful. The purpose of this paper is to give a general overview of some of the decisions which need to be made when devising a semantic differential scale.<sup>1</sup>*

## 1. Introduction

The semantic differential scale (SDS) is a scaling tool which has been used frequently for measuring social attitudes, particularly in the fields of linguistics and social psychology. It was first devised by Osgood, Suci and Tannenbaum (1957). Typically the scale is a seven-point bipolar rating scale using adjectival opposites, although some studies have used five- and six-point scales (these will be discussed below). The SDS has been used extensively in language attitude studies as a means of measuring subjects' attitudes towards various languages, dialects, accents, as well as the speakers of different varieties. Usually a tape is played of various speakers and subjects are asked to listen to the voices and make judgements about the speakers' personality, looks, etc. using an SDS. For example, a subject might be asked to rate a speech sample in terms of a number of personality features. A feature such as *honesty*, for example, would be represented by the semantic differential scale in the following form:

honest 1 2 3 4 5 6 7 dishonest

There are nearly always at least eight different pairs of adjectives, and quite often there are a dozen or more.

This scale has the advantage that the subjects can indicate whether they judge the speaker to be extremely honest or dishonest by marking the extremities (1 or 7, respectively), or whether they have not formed an opinion of the speaker's honesty by marking position 4, a neutral position half way between the two extremes. One advantage of this technique is that it forces subjects to focus on the expected dimensions since the categories are already provided (Agheyisi & Fishman 1970:148).

Another advantage is that it is relatively easily implemented. It is not a difficult concept for lay people to understand since essentially all you are asking them to do is to decide whether they think a speaker is or isn't x, or they don't know. Nor is the procedure complex since all that is required is for the subject to make some sort of judgement and circle a number.

---

<sup>1</sup> Many thanks to Dr. Kate Burrige for her helpful comments, suggestions and advice in the early drafts of this paper.

Nevertheless, the SDS has not been without its critics. Lee (1971), for example, argues that, on the whole, scales are not accompanied or supported by evidence of reliability or validity. Therefore it is not known, firstly, how much the listeners agree with each other, or, secondly, whether such a measure is reliable over time. However, the former problem can be addressed by using an SDS accompanied by some form of questionnaire or interview with the subjects to confirm its results. As for the latter, it seems that this is not so much a criticism specifically of the SDS, but of the methodology which has been employed in attitude studies generally. After all, any technique, whether it be an SDS, questionnaire, or interview can be repeated at a later time as a test—retest procedure. Lee's other main criticism is the *ad hoc* method researchers have used in selecting adjectives, without offering any empirical rationale to support their choices. Again, this is a criticism of research methodology rather than the scale itself. And, as we shall see below, there are procedures to follow which will ensure as much as possible that the measures are relevant to the subject group and aims of the research.

While the SDS may have shortcomings (and what system doesn't?), as we have seen above, its main drawbacks can be curtailed. Therefore the SDS still appears to be among the best available instruments for measuring multi-component concepts (Agheyisi & Fishman 1970) and has been employed very successfully in attitude studies. However, the purpose of this paper is not to discuss whether the use of the scale is valid or not, but to discuss what factors need to be considered when devising such scales since there are many ways of improving them. Much of the discussion will draw on my own experiences while conducting a pilot study which was part of a larger ongoing study on attitudes towards Australian accents. Examples will also be drawn from other attitude studies, both in Australia and overseas.

There are two main components of a semantic differential scale: (1) the adjectives; and (2) the layout of the scale itself. In designing a scale which can be used easily by subjects as well as providing meaningful results for the researcher, the researcher must make a number of decisions regarding both of these components. I will begin with the selection of adjectives since this is the most crucial aspect.

## **2. Selection of Adjectives**

In selecting adjectives, the researcher should not simply think up a range of adjectives he/she thinks might sound adequate and use them straight away in a study. There are certain procedures which should be followed to maximise the relevance of traits in order to make the data meaningful. It is standard and almost essential that a pilot study be carried out in order to ascertain which measures are most appropriate for the study at hand (Giles & Bourhis 1973:340). In his Chicago and Texas studies, for example, Williams (1974) examined the role that stereotypes play in the formation of Black and White teachers' attitudes towards children of different ethnicities. He outlined a four-step process which he used in selecting scales for these two studies (1974:23). The first step was to conduct a pilot study to elicit adjectives from a sample group. This sample group would be played a tape of speech samples and then be encouraged to discuss freely their impressions of the speakers they had just heard. He then devised a prototype SDS using the most common adjectives elicited from these discussions. This scale was then tested using another subject group. Finally, responses were quantified and inter-correlated in order to determine whether fewer or more basic dimensions could be identified so a final SDS could be prepared to be used in the main study. While this process is a good way of ensuring meaningful adjectives, unfortunately it does not usually elicit many antonyms for the most common (and relevant) adjectives. These are left for the researcher to devise.

## 2.1 Antonyms

A decision needs to be made regarding whether the antonyms are going to be complementary opposites (for example *honest—dishonest*) or whether they are going to be gradable and therefore more subtle, for example (*entertaining— boring*). At first glance using complementary antonyms seems more straight forward, and it *can* be a simple process for many adjectives, especially those on which a negative morpheme can be added (*successful— unsuccessful; honest— dishonest; patient— impatient; educated— uneducated; sincere— insincere; kind— unkind; reliable— unreliable*). Alternatively, the negator *not* can precede just about any adjective (*rich— not rich; good— not good; generous— not generous*), but this is less satisfactory because it does not really tell you very much. For example, *not rich* does not necessarily mean the same as *poor*. While the points on the scale allow for degrees of judgement, there is no way of telling whether the subject considers *not rich* to mean *poor* or perhaps of middle income.

For adjectives such as this gradable antonyms might be more meaningful. For example, *not generous* does not necessarily mean *stingy*, it could just mean *careful* or *thrifty*, and therefore *stingy* is more meaningful than *not generous*. Similarly, *sophisticated* is a less ambiguous antonym for *common* than *uncommon* (which can also mean 'rare'). This raises the very difficult problem of which antonym to choose because it is not always clear which adjectives are more appropriate. The best way to illustrate the complexities of making a decision is by giving some examples. The list below shows some alternatives for a small range of adjectives. All have either been found in attitude studies or were elicited as part of my own pilot study.

### Possible Antonyms

<b>serious</b>	good sense of humour, funny, cheerful, witty
<b>friendly</b>	unfriendly, unsociable, stand-offish, snobbish,
<b>stuck-up,</b>	distant, cold, shy
<b>unsuccessful</b>	successful, ambitious, satisfied
<b>likeable</b>	unpleasant, unpopular
<b>educated</b>	uneducated, unskilled
<b>intelligent</b>	unintelligent, slow, dull, dumb, uneducated
<b>unkind</b>	kind, warm-hearted, helpful, generous, easy going trustworthy untrustworthy, dishonest, two-faced, deceitful
<b>sincere</b>	insincere, dishonest, two-faced, deceitful
<b>easy-going</b>	fussy, perfectionist, highly strung, uptight,
<b>particular,</b>	fastidious

A person who is not friendly may be a snob, or shy, or neither. *Perfectionist* and *highly strung* are two antonyms for *easy-going*, yet the former is relating to individuals' actions, the latter to their temperament. *Dishonest*, *two-faced*, and *deceitful* are all antonyms for both *trustworthy* and *sincere*.

It can also be seen even from this small sample that some antonyms are more appropriate than others. For example *uneducated* is a more appropriate antonym for *educated* than *unskilled* because being uneducated does not necessarily mean being unskilled.

It can be seen that sometimes it is better to use complementary antonyms, and at other times gradable ones. However, there is no reason why a combination of complementary and gradable antonyms could not be used on the same scale. In fact it may be wiser to do so since both have their advantages and disadvantages, depending on the particular adjectives selected.

## 2.2 Homonyms and ambiguities

It is tempting to take adjectives from someone else's study thinking that since someone else has used them then they must be okay. However, as we shall see below, there can be pitfalls in doing this. Apart from eliciting adjectives, as part of my pilot study I decided to also test out some adjectives used in previous studies. One unforeseen complication arose with the adjective pair *intelligent—dull*. I found the feature *dull* was misinterpreted due to a voice quality of one speaker. Because this speaker had a quiet, monotone voice, it was suspected that at least one subject considered *dull* to mean 'boring' rather than 'unintelligent'. This was apparent since the majority of subjects rated this particular speaker as being somewhat intelligent, although quite unsocial.

Another point to note is that the meaning of the adjectives must be clear or they are likely to be misinterpreted by the subjects. The following is the list of adjectives used by Shuy and Williams (1973) which they selected from prior literature:

<i>dull—sharp</i>	<i>complex—simple</i>	<i>positive—negative</i>
<i>smart—dumb</i>	<i>worthless—valuable</i>	<i>difficult—easy</i>
<i>good—bad</i>	<i>careful—sloppy</i>	<i>weak—strong</i>
<i>thin—thick</i>	<i>rough—smooth</i>	<i>fast—slow</i>

It is difficult to imagine what is meant when a speaker is described as *thin* or *thick*,<sup>2</sup> or as *rough* or *smooth*. Even the *positive—negative* dimension is somewhat ambiguous. Does it equate to *optimistic—pessimistic* or as *confident—hesitant*? Is *dull—sharp* referring to intelligence, or wits, or perhaps to a 'sharp tongue'? Does *complex—simple* mean *sophisticated—common* or *fastidious—easy-going*? If it means the latter, what is the difference between *complex—simple* and *difficult—easy*? Are the dimensions of *weak—strong*, *fast—slow* referring to physical or mental traits? Is *smooth* a positive or negative trait (where it can have connotations of sleazy or sneaky)? Clearly there are a lot of ambiguities in this scale which means that the results it would provide would be almost impossible to interpret.

Sometimes an adjective may be used which is not so much ambiguous but may be simply uninterpretable. For example, at least three studies (Strongman & Woosley 1967; Giles, Baker & Fielding 1975; Ball 1983) have used the adjective *laissez-faire* as the opposite polarity to *ambitious*. Presumably subjects who do not speak French will at least understand that it means *not ambitious*, but surely an English adjective such as *indifferent* or *satisfied* would be more appropriate for English-speaking subjects? Of course the antonym chosen here depends whether you think *ambitious* is a positive or negative trait. In fact, the concept of positive and negative polarity is of crucial importance to SDS and will be discussed in some detail in section 3.1. For now, the above examples illustrate some problems which can arise as a result of using adjectives from a previous study.

Of course, another problem which may arise is that the adjectives may simply not be relevant for your particular subject group. The notion of relevance is discussed in the following section.

---

<sup>2</sup> In fact, Shuy and Williams (1973) were forced to eliminate *thin—thick* from their analysis due to its ambiguity

## 2.3 Relevance

### 2.3.1 Relevance to subjects

Deciding which adjectives are appropriate will depend upon the subject group, the aims of the study, and all the relevant social factors. Here it becomes even clearer why it may be more prudent to take the time and elicit adjectives openly from a sample group and test them in a pilot study as Williams (1974) suggested. It is important to have adjectives which will be both meaningful to the study at hand and have some relevance to the particular subject group.

Subject groups for the pilot study must be selected using the same criteria as the subjects to be used in the main study and the same stimulus speakers should also be used. This procedure is aimed at testing and eliminating any adjectives which may be inappropriate. Selecting groups according to the criteria to be used in the main study is important because what is relevant to one group might not be relevant to another.

For example, El-Dash and Tucker (1975) conducted their study in Egypt. Their subjects were mainly Egyptians, although there was also one American university group. They identified four main positive characteristics; intelligence, likeability, leadership, and religiousness. Religion plays a dominant role in daily life in Egypt and so is a very relevant feature for this study. However, *religiousness* has rarely been used in studies conducted in the West, although Gallois and Callan (1981) used it as part of their study in Queensland. In this case, however, speakers were a mixture of nationalities and cultures (Australian, British, Vietnamese, French, Greek and Italian). Speakers were likely to include Orthodox, Catholic, Church of England, Buddhist and atheist. The mixture in this study would suggest that *religiousness* is likely to be a relevant feature. *Religiousness* is not as likely to be relevant, however, if you are asking subjects to make judgements on working- and middle- class Anglo speakers in Australia since these speakers share the same nationality, ethnicity and culture, and religion is not a dominant social force in the way that it is in Egypt.

One other problem is that words can mean different things to different people. For example, *unfriendly* can be interpreted in many ways, as illustrated in the list above, of possible antonyms for *friendly*. Unfortunately, the only way of knowing exactly how the subjects perceived these words is to interview them afterwards (which is advisable to confirm the results of the SDS), but this may result in an unworkable number of categories. In the end, such fine distinctions are probably not important. Semantics is a slippery business and it is probably not worth pursuing unless an obvious ambiguity arises such as with an adjective like *dull*, as discussed earlier.

### 2.3.2 Relevance to study

In attitude studies the adjective pairs are broken down into underlying dimensions either by factor analysis (Williams 1974; Gallois & Callan 1981) or they can be selected *a priori* on the basis of previous research and sociolinguistic theory (Ryan 1979). There have been numerous ways of dividing features. The more popular ones have been *evaluation*, *potency*, and *activity* (Osgood *et al.* 1957); *competence*, *personal integrity* and *social attractiveness* (Lambert 1967); *power* and *solidarity* (Romaine 1980; Bayard 1990).

If the *a priori* method is chosen then care must be taken to ensure that the adjectives are truly representative of the dimensions selected. For example, the adjectives *beautiful—ugly* would be much more suited to the dimension of *social or physical attractiveness* than to the dimensions of *power* and *solidarity*.

## 2.4 Number of adjectives

There seems to have been no real correlation up to now between the numbers of speakers<sup>3</sup> the subjects are required to listen to and the number of traits they must decide on when marking an SDS. For example, Strongman and Woosley (1967) had four speakers and used 18 pairs of adjectives, and Seggie, Fulmizi and Stewart (1982) had only three speakers and used 23 traits. On the other hand, Alford and Strother (1990) had 12 speakers and used 24 traits, while Gallois and Callan (1981) also had 12 speakers but only used 16 traits.

While it is necessary to include enough adjectives to provide the researcher with the desired information, there should not be so many that the task becomes tedious for the subjects. The number chosen should therefore partly depend upon the number of voices the subjects are required to listen to. A large number of speakers and a large number of traits would mean that the subjects are much more likely to become bored than if one of these factors were reduced. The scale would be more effective with either fewer traits or fewer speakers. The subjects would feel less fatigued, and results would be more likely to be an accurate reflection of their true opinions, resulting in more meaningful data.

Of course this does not mean that if there are only two or three speakers then 40 or 50 traits can be used! It is unlikely that so many would be useful in an attitude study anyway. A little judgement and common sense must also prevail.

## 3. Layout of scale

### 3.1 Positive or negative polarity

While the procedure of filling in an SDS is fairly straight forward (although see El-Dash & Tucker 1975), careful thought about the arrangement of the polarity can make the task easier on the subjects. Once again, this is not necessarily a straight forward procedure. Three main areas of difficulty which can arise are discussed below.

(1) As illustrated earlier, one weakness of SDS is that it may not always be clear if a trait is positively or negatively valued. For example, Gaies and Beebe (1991:168) cite a code-switching study using the matched guise technique by Furuya-Nakajima and Vogt (1990) where *ambition* and *self-confidence* are considered to be negative traits by the Japanese, but are valued more highly by Westerners. This type of problem can arise even within a single society such as Australia, for example, *career oriented* may be viewed as a positive trait by some but a negative one by others. It may also be considered context-dependent and so people will rate it half way between the bi-polar adjectives, unable to make up their minds. For example, in some cultures it may be considered a positive feature for males but not for females.

(2) This example also illustrates the fact that adjectives can be gender- dependent. For example, an adjective such as *tough* may be viewed positively by males but negatively by females, regardless of the gender of the speaker, or it may be viewed negatively when judging females but not when judging males. It may also depend upon the social network of the individual. For example, Labov (New York, 1966) and Cheshire (Reading, 1982) both found that 'toughness' is considered to be a positive quality within the social networks of some teenage working class groups. The researcher has no objective way of knowing whether the attitude of the judge is positive or negative for such adjectives without some empirical knowledge of the social values of the group under investigation.

---

<sup>3</sup> Note that the number of speakers does not necessarily mean different speakers since the matched guise technique was used with most of the studies. However, subjects were unaware that the voices were the same speaker and judged them as though they were different speakers.

What is considered to be positive also depends upon the nature of the concept and the context. For example, Romaine recognised a similar problem with the features of *masculine—feminine* and noted that it is implicitly assumed that ‘it is positive for a man to sound masculine, but negative for a woman to sound masculine, and vice versa’ (1980:219). While this is indicative of the stereotype, there are exceptions. Therefore, sometimes context can dictate what is a ‘positive’ characteristic.

(3) Another difficulty is that some features do not necessarily have a negative/positive polarity. For example, features such as *active—passive* are not intrinsically positive or negative.

This brings us to the issue of whether to list all the positive features on one side and all the negative features on the other (Alford & Strother 1990), or to randomise the polarity for left-right position (Gallois & Callan 1981; Romaine 1980; Strongman & Woosley 1967). Of course it is easier for the subjects to associate the left side (for example) with positive traits and the right side with negative traits. However, the randomising method has the advantage that it does not bias the subjects’ opinions and adjectives such as *active—passive* are not forced into unnatural polarities. This method has therefore been the preferred choice in many attitude studies.

### 3.2 Five- six- seven- nine-point scales

Another decision which needs to be made is how many points of choice should be on the scale. Osgood *et al.* (1957) originally used a seven-point scale which allows for a finer grade of judgement than a five-point scale. It could be argued that a nine-point scale would be an even finer indicator of judgement, however it has rarely been employed in attitude studies (but see Giles, Baker & Fielding 1975). The most likely reasons for this are that, firstly, subjects would find it difficult to grade evaluations so finely, and secondly, the task would become tedious. Such a distraction would be counterproductive since the aim of the exercise is to get subjects to make a rapid assessment. This is because rapid assessments are thought to reflect a ‘truer’ evaluation in the sense that it gets to a person’s underlying beliefs and attitudes.

Some have used a five-point scale (Strongman & Woosley 1967; Bayard 1990), however they are almost as infrequent in attitude studies as nine-point scales. They do not grade as finely as a six- or seven-point scale and as a result may not be the best indicator of a subject’s opinion because there are essentially only three choices; extremely [feature], don’t know, and something somewhere in between. A finer grading would surely yield a more accurate picture about a subject’s attitudes towards a speech variety or the speaker of that variety.

This leaves us with six- or seven-point scales which have, in fact, been the most popular choices in attitude studies. I will begin with the six-point scale. Being an even number there is no *neutral* choice so it forces the subject to lean towards one evaluation or another. The danger here is that the researcher does not know whether a 3 or 4 circled on the scale means that the subject thinks the speaker is ‘a little bit’ [feature] or whether they were really neutral and were forced to make a choice. A neutral answer can mean that the subject is (a) undecided, (b) doesn’t think the feature has any social relevance, or (c) doesn’t think the feature can be ascertained from a voice sample. It may be that you *want* to force the issue, although when studying attitudes, a neutral answer can often tell the researcher just as much as a leaning one way or another. This choice will of course depend on the design and aims of the study being conducted. A seven-point scale, on the other hand, has the advantages of allowing neutrality and has enough gradation to give meaningful data, yet not be too tedious. I believe, therefore, that the seven-point scale is the best choice.

### 3.3 Circle numbers or tick lines

This may be a relatively minor point to consider, but choices here can make life a little easier on the subjects. Whether numbers or lines are used (see Fig. 1 below) may depend upon the scale's polarity. For example, if all the 'positive' attributes are listed on one side and all the 'negative' attributes are listed on the other, then numbers 1 and 7 are consistently associated with one polarity or the other and so the choice between numbers and lines will probably make no difference. On the other hand, if the polarity is randomised, numbers could be confusing as sometimes a 7 will refer to a 'positive' polarity and other times a 'negative' polarity. In this case lines might be more appropriate.

Figure 1: **Numbers vs Lines**

Numbers:	Educated	1	2	3	4	5	6	7	Uneducated
Lines:	Educated	—	—	—	—	—	—	—	Uneducated

### 4. Conclusion

It can be seen that there are many considerations at every step when devising a semantic differential scale. The decisions made can be crucial in eliciting data which is analysable, meaningful, and relevant to the study at hand. Although the choices made depend on the subjects and aims of the study, it has been shown that there are many ways to improve the quality of an SDS. Conducting pilot studies is crucial to both elicit and test adjectives. Antonyms (both gradable and complementary) should be selected according to those which are the most meaningful, appropriate, and relevant to the aims of the study and to the subjects. Homonyms and ambiguous antonyms should also be avoided. Finally, a combination of a seven-point scale, randomised polarity and lines instead of numbers on the scale also seem to be the best way of laying out the scale.

## REFERENCES

- Agheyisi, R. & J.A. Fishman. 1970. Language attitude studies: A brief survey of methodological approaches. *Anthropological Linguistics*. 12(5): 137- 157.
- Alford, R.L. & J.B. Strother. 1990. Attitudes of native and non-native speakers toward selected regional accents of U.S. English. *TESOL Quarterly*. 3(24):479-495.
- Ball, P. 1983. Stereotypes of Anglo-Saxon and Non-Anglo-Saxon accents: Some exploratory Australian studies with the matched guise technique. *Language Sciences*. 5(2):163-183.
- Bayard, D. 1990. 'God help us if we all sound like this': Attitudes to New Zealand and other English accents. In A. Bell & J. Holmes (eds), *New Zealand ways of speaking English*, 67-96. Clevedon: Multilingual Matters Ltd.
- Cheshire, J. 1982. Linguistic variation and social function. In S. Romaine (ed.), *Sociolinguistic variation in speech communities*, 153-166. London: Edward Arnold.
- El-Dash, L. & R. Tucker. 1975. Subjective reactions to various speech styles in Egypt. *International Journal of the Sociology of Language*. 6:33-54.
- Gaies, S.J. & J.D. Beebe 1991. The matched-guise technique for measuring attitudes and their implications for language education: A critical assessment. In E. Sadtano (ed.), *Language acquisition and the second/foreign language classroom*. Anthology Series 28. Singapore: SEAMEO Regional Language Centre.
- Gallois, C. & V.J. Callan. 1981. Personality impressions elicited by accented English speech. *Journal of Cross-Cultural Psychology*. 12(3):247-359.
- Giles, H. & R.Y. Bourhis. 1973. Dialect perception revisited. *Quarterly Journal of Speech*. 59(3):337-342.
- Giles, H., S. Baker, and G. Fielding, 1975. Communication length as a behavioural index of accent prejudice. *Linguistics*. 166:73-81.
- Labov, W. 1966. *The social stratification of English in New York city*. Washington DC: Center for Applied Linguistics.
- Lambert, WE. 1967. A social psychology of bilingualism. *Journal of Social Issues*. 23(2):91-109.
- Lee, R.R. 1971. Dialect perception: A critical review and re-evaluation. *Quarterly Journal of Speech*. 57:410-417.
- Osgood, C.E., G.J. Suci & P.H. Tannenbaum. 1957. *The measurement of meaning*. Urbana: University of Illinois Press.
- Romaine, S. 1980. Stylistic variation and evaluative reactions to speech: Problems in the investigation of linguistic attitudes in Scotland. *Language and Speech*. 23(3):213-232.
- Ryan, E.B. 1979. Why do low-prestige language varieties persist? In H. Giles and R. N. St. Clair (eds), *Language and social psychology*, 145-157. Oxford: Basil Blackwell.
- Seggie, I., Fulmizi, C. and Stewart, I. 1982. Evaluations of personality traits and employment suitability based on various Australian accents. *Australian Journal of Psychology*. 34(3):345-357.
- Shuy, R. and Williams, F. 1973. Stereotyped attitudes of selected English dialect communities. In R.W. Shuy & R.W. Fasold (eds), *Language attitudes: Current trends and prospects*, 85-96. Washington DC: Georgetown University Press.
- Strongman, K.T. and Woosley, J. 1967. Stereotyped reactions to regional accents. *British Journal of Social and Clinical Psychology*. 6:164-167.
- Williams, F. 1974. The Identification of linguistic attitudes. *International Journal of the Sociology of Language*. 3(1) :21-32.