

# The Linguistic Data Consortium Member Survey: Purpose, Execution and Results

Marian Reed, Denise DiPersio and Christopher Cieri

Linguistic Data Consortium, 3600 Market Street, Philadelphia, PA 19104 USA  
{mreed, dipersio, ccieri}@ldc.upenn.edu

## Abstract

The Linguistic Data Consortium (LDC) seeks to provide its members with quality linguistic resources and services. In order to pursue these ideals and to remain current, LDC monitors the needs and sentiments of its communities. One mechanism LDC uses to generate feedback on consortium and resource issues is the LDC Member Survey. The survey allows LDC Members and nonmembers to provide LDC with valuable insight into their own unique circumstances, their current and future data needs and their views on LDC's role in meeting them. When the 2006 Survey was found to be a useful tool for communicating with the Consortium membership, a 2007 Survey was organized and administered. As a result of the surveys, LDC has confirmed that it has made a positive impact on the community and has identified ways to improve the quality of service and the diversity of monthly offerings. Many respondents recommended ways to improve LDC's functions, ordering mechanism and webpage. Some of these comments have inspired changes to LDC's operation and strategy.

## 1. LDC Member Survey Goals

The External Relations Group at the Linguistic Data Consortium (LDC) is primarily concerned with the needs and satisfaction of LDC's user communities and extending LDC services to new organizations and communities. In order to best serve its members, LDC needs to remain aware of their changing needs and provide a mechanism for gathering and updating this information. The more LDC knows about its constituents communities, the better it can serve them and new communities not yet engaged.

### 1.1 LDC's Target Market

Language technology developers belong to a niche market, meaning that their needs are highly focused and cannot be satisfied by mainstream products. They include academic, government and commercial organizations whose research interests and performance goals vary widely within the sub-disciplines of text, speech, video and multimodal processing. As a result, there are few data collections that appeal to all of these communities. Finding a way to identify and address such highly differentiated needs is a challenge for the LDC. An ideal method for discovering market needs would have minimal administrative costs, be easily transmitted to a variety of end recipients and most importantly, be relevant to the communities involved.

Before 2006, LDC had no regular mechanism in place to elicit feedback from its research communities. LDC's management and External Relations group believed that members were satisfied with both the quality of data LDC released each year and the level of service provided, but there was no means to quantify members' sentiments. In 2006, the External Relations group decided to administer a survey to active participants in the LDC community, including members and nonmember organizations, in order to probe the following issues: the groups' awareness levels about LDC activities, their level of satisfaction with LDC data and services and their needs for the future. The remainder of this paper describes how

the 2006 and 2007 surveys were created, administered, evaluated and incorporated into LDC policy.

## 2. External Relations

LDC is a nonprofit, open, international consortium of universities, corporations, and government research organizations that creates and distributes linguistic resources, including data, tools and specifications. LDC also works with a variety of organizations in order to develop and disseminate best practices and standards for linguistic research. LDC was founded in 1992 with a seed grant from the Defense Advanced Research Projects Agency (DARPA). Ongoing operation costs are funded solely from membership and licensing fees while grants from commercial sponsors, other non-profits and the United States Departments of Commerce, Defense, Education and the Interior fund new resource creation. The University of Pennsylvania is LDC's host institution.

LDC's External Relations group consists of the following: the Communications Coordinator, who develops and executes new initiatives; the Membership Coordinator, who handles LDC's member and nonmember requests and administers data distributions for research projects and evaluations; the Publications Programmer, who produces LDC's general publications and limited releases for ongoing projects and evaluations; and the External Relations Manager, who oversees these functions and reports directly to the Executive Director.

The Communications Coordinator position was created in 2005 to demonstrate LDC's commitment to member satisfaction and to focus on identifying new areas for growth. In 2006, these efforts were still in the early stages of development and the Member Survey was an important tool in setting the groundwork for future activities.

### 2.1 External Communications

Generally, LDC focuses on strategies that solidify its membership base, create a strong sense of community within the Consortium and expand LDC awareness in and of related communities. Understanding the Human

Language Technology (HLT) communities and others involved in research in linguistics is a crucial part of that strategy because of how significantly the HLT field affects LDC's practices. LDC does not exist for its own purposes, but rather to serve these communities. By learning more about their members, LDC can better respond to and occasionally anticipate communal needs.

### 3. LDC's Membership Model

LDC's operations are dependent on two income streams: (1) revenue from membership and corpora sales; and (2) grant funds. Membership and licensing revenue supports LDC operations including corpus production, distribution and member services while grant funds are used for specific research-project goals. These areas are interrelated since specific project goals may require the enhancement of infrastructure and membership revenue may be used to extend LDC activities into new research domains.

Expanding the depth and breadth of its user communities allows LDC to facilitate the sharing of a greater variety of resources and to encourage interdisciplinary research. Memberships are important to LDC since they foster the consortium ideal and create a stronger sense of community. Memberships also help ensure that LDC can continue to make linguistic data relatively affordable and widely available for various research needs.

LDC currently offers the following membership options:

- Non-Profit/Government Standard Memberships
- Non-Profit/Government Subscription Memberships
- For-Profit Standard Memberships
- For-Profit Subscription Memberships
- LDC Online Memberships

Non-Profit/Government Memberships entitle members to use LDC data for language-related education research and technology development for non-commercial purposes. For-Profit members may use LDC data to develop and test products and systems, including commercial products. Standard Members receive rights to sixteen corpora from their membership year(s), and if they are members in the current year, they also receive free shipping and discounted fees on older LDC corpora. Subscription Members are entitled to two copies on media of all normal publications from the membership year as well as all other Standard Membership benefits. Subscription Members need not request monthly data, but receive it as it is published on or around the 15<sup>th</sup> of each month. LDC Online members do not have current membership rights to monthly publications as Standard and Subscription Members do. They simply have full access to LDC Online during the duration of their membership year.<sup>1</sup>

---

<sup>1</sup> LDC Online contains an indexed collection of Arabic, Chinese and English newswire text, English telephone speech from the Switchboard and Fisher collections and the American English Spoken Lexicon, as well as the Brown corpus. Members and guests may login at <https://online.ldc.upenn.edu/login.html>.

### 4. 2006 Member Survey

When carefully created, administered and analyzed, surveys are an effective marketing research tool (Bell, 1966). Email surveys are a low cost venture, and as a result, low response rates can be observed. However, surveys are user-friendly, can easily reach a wide audience and can be a relatively easy method of gathering primary market research. In addition, surveys give respondents an opportunity to comment on, and often directly affect, the proceedings of the organization administering the survey (Waddington, 2000).

LDC's External Relations group intended that the 2006 Member Survey would serve as a foundation for subsequent intra-community communications. The group was interested in collecting current member and user data as well as information about former LDC members who still licensed data intermittently. They also wanted to determine the current sentiments of Standard and Subscription Members, to discover why former members had not rejoined LDC and why some nonmember organizations have never joined LDC.

The Communications Coordinator decided that the survey should feature two distinct parts: (1) a general set of questions focusing on issues common to each group that would be administered to the entire recipient population and (2) a customized set of questions that pertained to the unique circumstances of each subset. Sequencing survey questions so that neutral questions are asked before more thought-provoking questions can elicit more survey responses (Marcus, 1975). Two principal criteria guided question development: first, what does the LDC need to know about its community and second, what questions will lead to the most relevant and greatest amount of feedback.

#### 4.1 2006 Survey Target

The target group for the 2006 Member Survey was identified as any individual who licensed data from the LDC or who joined as a member during 2005 and/or 2006. That group consisted of 919 individuals at 581 organizations. The target group was then divided into four subgroups:

- Current Standard Members (2005 or 2006)
- Current Subscription Members (2005 or 2006)
- Former LDC Members (members in any year from 1993 – 2004 who were not also current members)
- Nonmembers (individuals who had never been members)

#### 4.2 2006 Survey Structure

The survey ultimately consisted of two parts. The first part was the twelve question general survey that was sent to the entire targeted population. Questions centered on common experiences within the group and included the following: the current LDC Catalog (both the corpora themselves and the Catalog's search function), the LDC homepage, the desirability of n-gram corpora and a rating of each individual's interactions with the Membership Department. The second part consisted of four discrete surveys (4-5 questions each) specifically targeted to the subgroups:

- Standard Members were asked if they were satisfied with their memberships and if they had considered joining as Subscription Members
- Subscription Members were asked about their opinions on, and satisfaction with, various elements of the Subscription Membership
- Former Members were asked why they did not renew their memberships
- Nonmembers were asked if they knew about LDC Memberships and the rights they offered

### 4.3 Administering the 2006 Survey

The survey was sent by email on 13 November 2006 and recipients were given two weeks to reply. Fifty individuals replied within this initial period. On 19 December 2006, a reminder email was sent which resulted in an additional 29 responses. Survey responses were saved in a database that was accessible through the LDC intranet to administrators. Results were tallied for the group as a whole and separately for each subgroup. Each individual reply was also accessible in a separate table that listed the date and time of each response and the name and affiliation of the respondent.

## 5. 2006 Survey Analysis

Most survey responses are received within 24 hours of its distribution and nearly all replies are received within two weeks (Hamilton, 2003)<sup>2</sup>. This was true in the 2006 Member Survey; the majority of initial responses (88.5%) were received within a fortnight of the original sent date. The reminder email, sent in December, was deemed necessary because of the survey's timing (proximity to global holidays) and concern over sensitive spam filters. The reminder email generated over half as many responses as the original survey email.

LDC intended that the Survey reach multiple persons at a given organization, not just the primary contact, in order to increase the chances of a reply from each targeted organization. In other words, LDC was more interested in receiving a reply from each surveyed organization than each surveyed individual<sup>3</sup>. LDC received 79 unique responses (out of 919 emails sent) for an overall response rate of 8.6% (not taking into account the number of bounce-backs and failed email attempts). However, nine surveyed companies and universities had duplicate replies, so the organizational response rate drops to 70 unique surveys. Seventy replies from a pool of 581 organizations yielded a response rate of 12.1%, above the 10% threshold generally considered the mark of a successful email survey (Stewart, 2007).

LDC received a greater percentage of survey responses from current members than from former members and nonmembers. Because 2005 and 2006 members have a more current relationship with LDC, it can be assumed that they would be more interested in responding to a survey that would have an impact in their

immediate community. Former and nonmembers may have a lower level of interest in the Consortium since they are not actively participating in it (Smith, 2005). Current members have the greatest amount of contact with the LDC through their current memberships, though they represent a smaller group than Former and nonmembers.

Below are the response rates per group. Current members are listed as 'higher interest recipients' (based upon their higher level of involvement with LDC) and all non-current members are identified as 'lower interest recipients':

#### Higher interest recipient response rates

- Current Standard Members – 36 replies, 24.8% of group
- Current Subscription Members – 17 replies, 42.5% of group

#### Lower interest recipient response rates

- Former LDC Members – 2 replies, 2.4% of group
- Nonmembers – 24 replies, 7.4% of group

Some survey responses warranted individual follow-ups and the External Relations group conducted them in the first quarter of 2007<sup>4</sup>.

### 5.1 2006 Survey Conclusions

The 2006 Survey showed that LDC has an extremely high approval rating across the board, specifically with its current members (94% very favorable or better). A variety of corpora were deemed the most useful and many suggestions for future publications were received. A few respondents provided constructive feedback on individual corpora, the usability of the LDC Catalog and the sixteen-corpus limit for Standard Members.

LDC has benefited directly and indirectly from the 2006 Survey. First, LDC gained a better understanding of the needs of its communities and created an open dialogue through which its members can express themselves to the Consortium. Also, certain responses provided LDC with suggestions for ways to improve its internal procedures, suggestions that LDC will consider as it develops future policies. Indirectly, LDC believes that the Survey created goodwill in the community by providing recipients with a greater sense of inclusion in the Consortium.

The surveyed communities also directly benefited from their efforts. Once all the responses were received, LDC emailed a synopsis of the survey results and selected comments to the entire target population in early 2007. This allowed survey respondents to see how their unique circumstances compared to others organizations and potentially provided a clearer understanding of LDC's community to each survey recipient.

It has been concluded that the 2006 Survey was a success. LDC received valuable input from its most interested users and greater insight into member perspectives. The survey was also relatively inexpensive to administer, analyze and document.

<sup>2</sup> Hamilton noted that 96.5% of responses are generally received within two weeks of the initial survey.

<sup>3</sup> Some contacts at member organizations are librarians and billing personnel not necessarily familiar with LDC data sets. Their responses would therefore not be representative of target users.

<sup>4</sup> Some survey respondents specifically requested follow-up emails. A small percentage of respondents left vague replies that needed to be clarified while a handful of others had ideas that warranted the creation of a dialogue.

## 6. 2007 Member Survey

In order to build on the baseline established by the 2006 Member Survey, LDC decided that a follow-up 2007 Survey should be conducted.

### 6.1 2007 Survey Target

The 2007 Member Survey was administered to the same target group, but was updated to reflect then current members (2006 and 2007 members) as well as all parties invoiced during 2006 and 2007 (as opposed to the 2005-2006 pool targeted in the 2006 Survey). However, there was another important difference between the two surveys. In 2007, LDC offered a \$500 benefit that could be applied toward a membership or corpus purchase in 2008 to one blindly-selected winner. The Survey cut-off date was 14 January 2008 and any respondent who had submitted a reply before then was eligible to receive the benefit. It has been shown that offering an incentive has affected survey response rates (Ryu, 2005) and LDC was interested to see if an incentive could increase the number of participants in the 2007 Survey.

### 6.2 2007 Survey Structure

The 2007 Survey was slightly reworked to reflect current areas of interest, though in order to establish a basis for comparing the two surveys, and to confirm that the 2006 Survey results were valid, the 2007 Survey repeated many questions from the previous year. The basic survey structure was unchanged; every recipient received the 14-question general survey portion and each sub-group received a specifically tailored subset of 4-5 questions. General questions again focused on common experiences within the group and included the following topics: a rating of the LDC Catalog (both content and the Catalog's search function) and the LDC homepage and a rating of the group's experiences with the Membership Department.<sup>5</sup>

New questions in the 2007 Survey concerned the feasibility of a web-based LDC Suggestion Board and preferences for particular high-density DVD publication formats. Sub-group questions were largely unchanged, though each 2007 Survey included a text box allowing recipients to directly comment on how LDC could improve or enhance their membership experiences.

### 6.3 Administering the 2007 Survey

The 2007 Member Survey was sent to 1730 individuals at 1021 organizations, nearly double the number of those who received the 2006 Survey. The survey was first sent via email on 17 December 2007. Due to proximity to the holidays, a reminder email was again deemed necessary. The first email generated 48 replies, a significantly lower initial response rate than seen in 2006. However, the reminder email sent on 7 January 2008 generated 189 additional replies for a total of 237 total replies. Thirty-three replies were from intra-organizational duplicates resulting in 204 unique surveys.

<sup>5</sup> Due to the overwhelming interest in n-gram corpora and the fact that an n-gram corpus is not on LDC's 2008 publications schedule, the question inquiring about the desirability of an n-gram corpus was omitted from the 2007 Survey.

Therefore, the 2007 Survey yielded a 19.9% organization-based response rate<sup>6</sup>, over one and a half times the rate in 2006 (12.1%).

## 7. 2007 Survey Analysis and Comparison to 2006 Results

The increased response rate for the 2007 Survey was reflected in each of the subgroups:

### Higher interest recipient response rates

- Current Standard Members – 73 replies, 68.9% of group (2006 – 24.8%)
- Current Subscription Members – 23 replies, 52.3% of group (2006 – 42.5%)

### Lower interest recipient response rates

- Former LDC Members – 17 replies, 7.9% of group (2006 – 2.4%)
- Nonmembers – 92 replies, 13.5% of group (2006 – 7.4%)

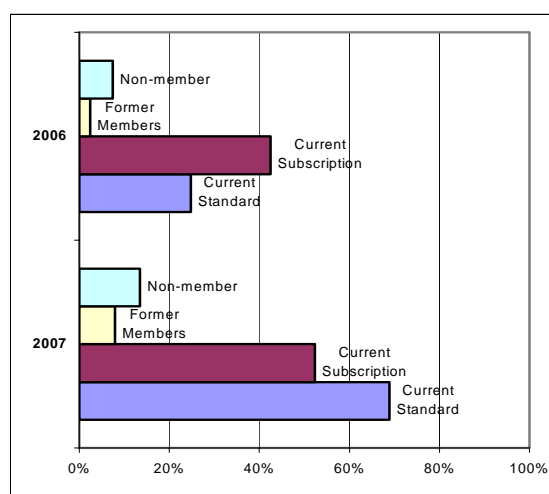


Table 1: Comparison of survey responses by year

There are two principal explanations for the increased response rate in 2007. First, the incentive (\$500 benefit) may have stimulated more people to respond. The increased response rates in the 'lower interest recipient' pool supports this conclusion since former members and nonmembers pay licensing fees for each corpus they acquire and the financial boon may have appealed to their needs.

Second, respondents received the January reminder email after the holidays and when they were actually able to access their professional email accounts. In 2006, the first and second notifications were sent in November and December, respectively, when year-end concerns were likely paramount and when many people were on extended holidays.

The 2007 Survey confirmed the high satisfaction levels expressed in 2006, with 96% of respondents being satisfied with the LDC and its Membership services. Due to the higher number of responses, more specific follow-ups were needed in early 2008 than in the previous year

<sup>6</sup> This figure does not take into account the number of bounce-backs and failed email attempts for the 2007 Member Survey.

and LDC reviewed a proportionally greater number of data suggestions as well.

## 8. Changes to LDC Policy Resulting from Member Surveys

As a direct result of the Member Surveys, many changes have been made to the LDC Catalog, the LDC home page and subtle adjustments have been made to LDC's practices. For example, survey respondents expressed some frustration regarding the searching the LDC Catalog. Users described difficulties while searching for corpora that they know exist as well as while attempting to search for a corpus that would meet their research and developmental needs. As a direct result of these comments, LDC is investigating ways to improve and update the Catalog search mechanism. In response to comments about unclear information on the LDC Homepage and FAQ, LDC modified multiple sections of its website. Although no formal policy changes have been made to date as a direct result of the two surveys, LDC's increased awareness of the community's needs may have a great impact on future policy decisions. Survey results have been documented and are easy for LDC employees to access, making certain that the surveys' findings can be referenced as needed.

## 9. Conclusion

The 2006 and 2007 Member Surveys were successful. Previously unconfirmed member opinions were validated, documented and quantified and the sense of the LDC community continues to be strong. Membership and corpora revenue grew in 2007, and some of that growth may be attributed to the knowledge gained, the goodwill created, and the changes made as a result of the Member Surveys. LDC's methodology has produced telling results that help gauge the subtleties of a sophisticated, complex niche market. By opening an additional means of communication, LDC also increases its chances of increasing member satisfaction in the future (Waddington, 2000).

LDC believes that administering additional Member Surveys in the future will be beneficial to the organization and to the community. Future surveys will not necessarily occur annually, but on an as-needed basis to address particular issues.

In regards to the timing of an 'annual' survey, it would appear that administering a survey earlier in the year will generate more replies than one administered during the closing months of the year. The chance to receive the \$500 benefit played a role in increasing the number of survey replies, but the logical conclusion is that more people replied in January because they were in fact checking their email and there was no general trend of absenteeism as a result of the year-end holidays (MarketTools, 2005).

The 2007 Survey results are available on LDC's homepage, [www ldc.upenn.edu](http://www ldc.upenn.edu)<sup>7</sup> to help encourage

ongoing discussion of issues related to resource distribution.

## 10. References

- Bell, Martin L. Marketing: Concepts and Strategy. Boston: Houghton Mifflin, 1966.
- Hamilton, Michael Braun. "Online Survey Response Rates and Times: Background and Guidance for Industry." 2003. Tercent, Inc./SuperSurvey. 24 March 2008  
<[http://www.supersurvey.com/papers/supersurvey\\_white\\_paper\\_response\\_rates.pdf](http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.pdf)>.
- Heskett, James L. Marketing. New York: Macmillan Publishing Co., Inc., 1976.
- "Holiday vs. Non-Holiday Research Results." December 2005. MarketTools. 24 March 2008  
<[http://www.markettools.eu/resources/files/holiday\\_research\\_brief.pdf](http://www.markettools.eu/resources/files/holiday_research_brief.pdf)>
- Marcus, Burton, David Aaker, and Dorothy Cohen. Modern Marketing. New York: Random House, 1975.
- Ryu, Erica et al. "Survey Incentives: Cash vs. In-kind; Face-to-Face vs. Mail; Response Rate vs. Non-response Error." International Journal of Public Opinion Research 18.1 (1 July 2005): 89-106.
- Smith, J. Walker, Ann Clurman, and Craig Wood. Coming to Concurrence: Addressable Attitudes and the New Model for Marketing Productivity. Evanston, IL: Racom Communications, 2005.
- Stewart, Morgan. "2005 Email Marketing Response Rate Study." Exact Target. 2007. 24 March 2008  
<<http://email.exacttarget.com/ETWeb/resources/pdf/2458.pdf>>.
- Waddington, Heather. "Types of Survey Questions." Encyclopedia of Educational Technology. 2000. 24 March 2008  
<<http://coe.sdsu.edu/eet/Artciles/surveyquest/index.htm>>

---

<sup>7</sup> 2007 Survey results are publicly available at [https://secure ldc.upenn.edu/intranet/surveyStatsPublic\\_2007.jsp?survey\\_id=1](https://secure ldc.upenn.edu/intranet/surveyStatsPublic_2007.jsp?survey_id=1)